

STIC-Biotech/ChemLib

58562

From: Davis, Natalie
Sent: Thursday, January 17, 2002 2:55 PM
To: STIC-Biotech/ChemLib
Subject: 09/823101

Please search for the protein encoded by SEQ ID NO: 1-4

Thanks
Natalie A. Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

RECEIVED
JAN 17 2002
STIC

Searcher: B. K. - Early
Phone: 308-4501
Location: Biotech Lab.
Date Picked Up: 1/17/02
Date Completed: 1/18/02
Searcher Prep/Review: _____
Clerical: 6 min
Online time: 3 min

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 4
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ABSS02
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)


```
|||||
75 ProProProProLeuProProSerPhePheLeuProProProLeuSe 91
546 .....
91 rProLeuLeuLeuLeuPheLeuProPhePheLeuLeuLeuPheP 108
545 .....TTCCTTAATTTTGTTCCTCTACTCTCT 516
108 heProLeuLeuProLeuPhePheLeuPheLeuLeuLeuPro 124
515 TATTCCTCC.....CCCTTTAAATAA 493
125 SerCysSerSerSerSerLeuLeuLeuLeuPheLeuProPhe...SerSe 140
492 ATTACTCTCTCTTTTCATATTGTATCTCCCTCCGCTAATAAAAA 443
140 rLeuSerSerSerProSerPheSerPheSerPheSerProSerSers 157
442 AAATAACACCGCATGTGTTTGTATTCTGTTTCGTTTATCCATC 393
157 erPheProPheLeuLeuProPhePheLeuLeuSerLeuPheLeu 173
392 TTCCTGTGATATTTCCTGTCGCGCTCTCCGCCCGCCCATAT 343
174 PheLeuLeuSerLeuLeuCysPheSerLeuLeuProProProPheph 190
342 AAATATCTCTGTTATTA 324
190 eLeuLeuLeuLeuLeu 196

seq_name: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:AA28950
seq_documentation_block:
ID AA28950 standard; Protein; 242 AA.
XX
XX AA28950;
XX
XX 17-OCT-2001 (first entry)
XX
DE Peptide #2987 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX
XX genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX MPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 29219; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:
XX
```

```
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 242 AA;
SQ

alignment_scores:
Quality: 92.50 Length: 123
Ratio: 1.745 Gaps:
Percent Similarity: 43.089 Percent Identity: 26.829

alignment_block:
US-09-823-101-3/rev x AA28950 ..

Align seg 1/1 to: AA28950 from: 1 to: 242

578 CCTCCTCCCGCCCGCAGCAAAATAAGTTTGTG..... 546
|||||
75 ProProProProLeuProProSerPhePheLeuProProProLeuSe 91
546 .....
91 rProLeuLeuLeuLeuPheLeuProPhePheLeuLeuLeuPheP 108
545 .....TTCCTTAATTTTGTTCCTCTACTCTCT 516
108 heProLeuLeuProLeuPhePheLeuPheLeuLeuLeuPro 124
515 TATTCCTCC.....CCCTTTAAATAA 493
125 SerCysSerSerSerSerLeuLeuLeuLeuPheLeuProPhe...SerSe 140
492 ATTACTCTCTCTTTTCATATTGTATCTCCCTCCGCTAATAAAAA 443
140 rLeuSerSerSerProSerPheSerPheSerPheSerProSerSers 157
442 AAATAACACCGCATGTGTTTGTATTCTGTTTCGTTTATCCATC 393
157 erPheProPheLeuLeuProPhePheLeuLeuSerLeuPheLeu 173
392 TTCCTGTGATATTTCCTGTCGCGCTCTCCGCCCGCCCATAT 343
174 PheLeuLeuSerLeuLeuCysPheSerLeuLeuProProProPheph 190
342 AAATATCTCTGTTATTA 324
190 eLeuLeuLeuLeuLeu 196

seq_name: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:AA04184
seq_documentation_block:
ID AA04184 standard; Protein; 242 AA.
XX
XX AA04184;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #2866 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
```



```

72 scys..... 73
    |||
161 TTCACGCAGCACTGCTGTCTGCATCATATGACTGTTGAGAATTATAGC   112
      |||| |  ||||||:::  |||||  ||
74 .... ArgMetcysGlnTyValIleGlucArgAlaGIu AlaAlaGI 88
                                |||
111 ATACCGTAGAGCACGATGCATGATCCATGGTCCACAGACAAGAATTCACGCCA   62
     |||  |||||  |||||  ::|||  ::  :::::::::::|||
88 yTYrPheALgThPhetyrGLyLysArGPHeAsnPHeGiNGluPrOGYL 105
        |||  ||||:  |||  |||  |||  ::|||:::
61 AAGCACACCTGTGTGACAGSCATAGGCTCCACAGACGAACCAGTGAGACA   12
          |||  |||:::  |||  |||  |||  ::|||:::
105 YSTYrVallLeuAlaARGly.....ThrLySGlYclYasprTrpsEr 118

seq_name= /SIDB8/gcgdata/geneseq/genesep/AAM98.DAT:AAM59825
seq_documentation_block:
ID       AAM59825 standard; Protein; 555 AA.
XX
XX       AAM59825:
XX
XX       07-DEC-1998 (first entry)
XX
XX       Cyridina luciferase.
DE
XX       Luciferase; reporter; analysis; exocytosis;
KW       cellular microenvironment; optical label.
OS       Cyridina hllegendorffii.
XX
XX       W09836081-AZ.
FN
XX       20-AUG-1998.
PD
XX       13-FEB-1998; 98WO-USO2774.
PF
XX       14-FEB-1997; 97US-0036805.
PR       13-FEB-1997; 97US-0038179.
XN
XX       (SLOK ) SLOAN KETTERING INST CANCER RES.
PA
XX       De Angelis DA, Miesenbock G, Rothman JE;
PI
XX       WPf: 1998-457118/39.
PT
XX       N-PDB; AAV41716.
DR
XX
XX       New polypeptide for monitoring changes in molecular environment,
PT especially release of synaptic vesicles - comprises targeting
PT component and reporter producing optical signal in contact with
PT second compartment, also new pH-sensitive mutants of green
PT fluorescent protein and related nucleic acid

Claim 10; Fig 13; 127pp; English.
XX
XX This is the amino acid sequence of Cyridina hllegendorffi
XX luciferase. The invention relates to methods and compositions
XX which utilise the emission of light to monitor changes in cellular
XX microenvironments. It is especially useful for monitoring
XX exocytic activity such as quantal release of synaptic vesicles.
XX Fusion proteins of Cyridina luciferase and synaptotagmin-1 or
XX vAMP/synaptobrevin-2 were targeted to synaptic vesicles and, upon
XX exocytosis, formed light-emitting complexes with luciferin present
XX in the extracellular medium. Photon emissions in the presence of
XX a depolarising stimulus can be observed with these systems. The
XX invention allows analysis of cells, cultures, tissue sections etc.
XX Sequence 555 AA;

```

	Percent Similarity:	38.060	Percent Identity:	28.358
alignment_block:				
US-09-823-101-3/rev x AAM59825	..			
Align seg 1/1	to: AAM59825	from: 1	to: 555	
372	TGTCGGCTCTCTCCCCCCCCTCATATATAATCTGTATTAA	323		
	::::			
22	CysProvalGlualagluprofer.....	30		
322	TGTGGGGGTGAATAATTTTGGGGGAAAACATCATGCCCCCATCG	273		
31SerThrProthrValprothrserc	39		
272	GGATAGCTCCNCAC.....ANGTCAGCTTGCTATA.....	241		
	:::			
39	ygluAlalalsolucllylucystlleasphrrargcysalaThCyls	55		
240CTCTGTCAAACCAANTGTGGACACACAAG	212		
56	ArgaspileleuserApsglyLeucysgluanslnysProglyLysrhc	72		
211	TGTGTCTCATGTGTGTTGACACACGTCATGTAAGTGCATGCACACG	162		
72	sCys.....	73		
161	TTCCACGACACTTGTCTGTGCATCGATATGACTGTGACATTTATAG	112		
74ArgmetCysglnInryValllegLucySarqvalglu.Alaalagl	88		
111	ATACGCTTAGGACAGATGATAGCACAGACTCACAGAAGAAGATCCAGCA	62		
88	tyrPhearThrPhetyrgLyLysargPheasnPheglInlurproLYL	105		
61	AAGCACACCTGCTGCACAGGCATAGGGTCCACAGACAGAACCCAGTGCACA	12		
105	ystryValleuAlarlgly.....ThrlysGlyIyAsprTPsr	118		
seq_name: /SIDUS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AA51162				
seq_documentation_block:				
ID	AA51162	standard;	Protein;	555 AA.
XX	AC			
XX	AA51162;			
XX	DT			
XX	21-MAR-2001	(first entry)		
XX	DE			
KM	Vargula (Cypridina) luciferase protein sequence SEQ ID NO:4.			
KW	Bioluminescence generating system; fluorescent protein; recreation;			
KW	entertainment; amusement; toy; paint; slimy play material; textile;			
KW	bubble making toy; balloon; cosmetic; bath powder; body lotion; gel;			
KW	powder; cream; toothpaste; dentifrice; soap; body paint; bubble bath;			
KW	gelatin; icing; frosting; beverage; beer; wine; champagne; soft drink;			
KW	glowing ice; fountain; liquid firework.			
XX	OS			
XX	Vargula hilgendorffi.			
XX	PN			
XX	US6152358-A.			
XX	PD			
XX	28-NOV-2000.			
XX	PE			
XX	17-AUG-1998;	98US-0135988.		
XX	PR			
XX	06-FEB-1996;	96US-0597274.		
XX	PR			
XX	25-NOV-1996;	96US-0757046.		
PA	(BRYA/) BRYAN B.			
XX	PI			
XX	Bryan B;			
XX	WP1; 2001-060056/07.			

DR N-PSDB: AAC92232.
XX
PT New combinations of articles of manufacture and with bioluminescence
PT generating systems, useful for producing novelty items for
PT entertainment, recreation and amusement, personal items or foods, e.g.
PT toys, cosmetics or gelatins
XX
PS Disclosure: Column 105-110; 83pp; English.
XX
CC The present invention describes a combination comprising an article of
CC manufacture and one or more components of a bioluminescence generating
CC system. The combination is a novelty item and the article of manufacture
CC is a paper product. The combination is useful for producing novelty
CC items, particularly those that glow, or produce or expel a
CC bioluminescent composition. The novelty items are useful for
CC entertainment, recreation and amusement, personal items, or foods.
CC Specifically, these include toys, paints, slimy play materials, textiles
CC (e.g. clothing), bubbles in bubble making toys and other toys that
CC produce bubbles, balloons, cosmetics, bath powders, body lotions, gels,
CC powders and creams, toothpastes and other dentifrices, soaps, body
CC paints, or bubble baths, gelatins, icings and frostings, beverages (beer,
CC wine, champagne, or soft drinks), glowing ice, fountains (e.g. liquid
CC fireworks) and other jets or sprays or aerosols of compositions that are
CC solutions, mixtures, suspensions, powders, pastes, or particles. AAC92229
CC to AAC92241 and AAB51159 to AAB51172 represent sequences which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 555 AA;
XX
Alignment_scores:
Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6
Percent Similarity: 38.060 Percent Identity: 28.358
XX
Alignment_block:
US-09-823-101-3/rev x AAB51162 ..
Align seg 1/1 to: AAB51162 from: 1 to: 555
372 TGTCCGCTCTCTCCCCCCCTCATATAAATATCTTGTATTAA 323
|||||
22 CysProValGluAlaGluProProSer..... 30
322 TTGGTGGGGGTGATTAATTTTGGGGGAAAACTCATGTCCTCCCATCGG 273
31SerThrProThrValProThrSerC 39
272 GGATAGCTCCNCAC.....ANGTCAGTTGTGCTATA..... 241
|||
39 ysglualalaysgluylucysileaspThrArgCysalathrcyslys 55
240CTCTGTTCAAAACANTGTGGACACACAG 212
56 ArgaspIleleuserAspSclYleucysGluAsnLysProGlyLysThrCy 72
211 TTGTGCTCATGTGGTGGACACGTCATCGTAACATGCATGCCACAG 162
|||
72 scYs..... 73
161 TTCACGACGACGCTTGTCTGTCATGATGATGACTGTGAGATTATAGG 112
||| ||| |||
74ArgmetCysgluInyValIleGluCysArgValGlu.AlalaIel 88
111 ATACGCTAGACCAGAGATGCATAGACAGATCACAAGAGATCCAGGCA 62
||| ||| |||
88 ylyrIphearGthrPheTyrglyLysrPheasnIpeGluInuProGlyL 105
61 AAGCACACCTGGTGACAGCATAGGTCACACAGAGACGACGACGAGACA 12
|| |||
105 ystTyValleuAlaargly.....ThrLyselGlyLyspripSer 118
seq_name: /SID58/ycgdata/geneseq/AA1988.DAT: AAP80656

seq_documentation_block:
ID AAP80656 standard; protein: 562 AA.
XX
AC AAP80656:
XX
DT 07-OCr-1990 (first entry)
XX
DE Tissue plasminogen activator analogue.
XX
KW Tissue plasminogen activator; analogue; K1 domain.
XX
PN EP293934-A.
PD 07-DEC-1988.
PF 03-JUN-1988; 88EP-0108949.
PR 04-JUN-1987; 87US-0058217.
PA (EISA-) EISAI CO. LTD.
PA (ZYMO-) ZYMOGENETICS INC.
PA (NOVO-) NOVO INDUSTRI A/S.
PI Mulvihill ER, Nexo BA, Yoshitake S, Ikeda A, Suzuki S, Hashimoto A;
PI yazuritha T;
XX
DR WPI: 1988-347625/49.
DR N-PSDB: AAN81091.
XX
PT New tissue plasminogen activator analogues with non-native K1-domain -
PT having higher fibrin specificity, and new encoding DNA sequences.
XX
PS Disclosure: 4 App; English.
XX
CC The polypeptide is encoded by a mutated t-PA sequence in plasmid
CC Zem99-9200. The t-PA analogue has the K1 domain of native t-PA replaced
CC by another kringle domain, esp. the K1 domain of plasminogen, which
CC mediates the binding of the analogue to fibrin and contains 6 Cys.
CC The analogue has greater specificity for fibrin than native t-PA. It is a
CC useful fibrinolytic agent with specificity for clot lysis. The analogue
CC can be tailored to increase plasma half life, solubility etc. It may also
CC include a finger, or growth factor domain of native t-PA, protein C
CC or factor VII.
XX
SQ Sequence 562 AA;
XX
Alignment_scores:
Quality: 83.50 Length: 171
Ratio: 1.210 Gaps: 11
Percent Similarity: 40.351 Percent Identity: 24.561
XX
Alignment_block:
US-09-823-101-3 x AAP80656 ..
Align seg 1/1 to: AAP80656 from: 1 to: 562
40 ATGCCGTGACACGAGTGTGCTTGGCT.....GGATCTTCTTG 77
:::||||| |||
81 ValProValLysSerCysSerGluProThrCysPheasnGlyGlyThrCy 97
78 TGAGTCTCGCTA..... 90
1:::|
97 scInuAlaIaleuTyrrPheSerAspPheValysgluInuProGluGlyP 114
91TGCAATCTGTGCTAGCGTATCTATATATCTCAACAGT 129
|||
114 heaIaglyLysCys.....Ser 119
130 CATATGATGACACAGACAGCAAGTGGCGTGAACGCT.....G 167
:::||||| |||
120 GluIleasp.Thr.....ArgAlaThrCysTyrgluAspGing 132

```

168 GCATGACGTTGACGATGAGTGTGTCACCAACATGAGACCAACTTGTG 217
    |||:|||||:|||||:|||||
132 LyleSerTyArgGlyThrTrpSerThrIaGluArgAla...AlaGlu 147
    |||:|||||:|||||:|||||
218 TGTCACACANTGTTGAACAGAGTAGACACAGCTGACNTGTGNGAGCC 267
    |||:|||||:|||||:|||||
148 CysThrAsnTrpAsnSerSerAlaLeuAlaGlnLysProTyrSerGlyAr 164
    |||:|||||:|||||:|||||
268 TATCCCGATGGGGGACATGATTTTCCCAAAAATTTATCCACCCC 317
    |||:|||||:|||||:|||||
164 GARGProAspAla..... 168
318 ACCAATTAAATACAGAGATTTTATATGAGGGGGGAGAGAGACG 367
    |||:|||||:|||||:|||||
169 .....IleArGLeuGlyLeuGlyAsnHisAsn 177
368 GGACACAGGAATATCAACAGAGATGCTAAACGAAAACAGAAATAC 417
    |||:|||||:|||||:|||||
178 AspCysArg.....AsnProAspArgAspSe 186
418 AAAAACAACATGCGGTATTTTATTTTATAGGACAGAGAGAT 467
    |||:|||||:|||||:|||||
186 rLys.....AlaTrpCysTyrValPhe.....LysGlyGlyLysT 198
468 ACAATATATGAA 478
    |||:|||||:|||||:|||||
198 YrSerSerGlu 201

```

seq_name: /stds/gcgdata/geneseq/geneseqp/AA2001.DAT:AA672099

seq_documentation_block:

ID AAG72099 standard; Protein; 328 AA.

AC AAG72099;

DT 31-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1780.

KM Human: olfactory receptor; OR: primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

DR WPI; 2001-290713/30.

PS Claim 11; Page 1169-1170; 1857pp; English.

CC The present sequence is an olfactory receptor which is encoded by
 CC one of a number of novel polynucleotides. The polynucleotides can be
 CC used in screening for olfactory agonists and antagonists. The methods
 CC allow for the determination of primary scents and the identification
 CC of the odour receptors used to detect these primary scents. The methods
 CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such
 CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.

SO Sequence 328 AA;

alignment_scores:

Quality:	83.00	Length:	196
Ratio:	0.933	Gaps:	8
Percent Similarity:	45.408	Percent Identity:	24.490

alignment_block:

US-09-823-101-3/rev x AAG72099 ..

Align seq 1/1 to: AAG72099 from: 1 to: 328

```

680 TATTTTGTTCATATAGTCGCCCCCACTAATATTTCTCTCG 631
    |||:|||||:|||||:|||||
28 PhePheLeuPheLeuValIleTyrValValThrLeuLeuGlyAsnSerIst 44
    |||:|||||:|||||:|||||
630 CGTCCCTCCATAGCTGCATATATATATATATTTCTCTCTCG 581
    |||:|||||:|||||:|||||
44 yLeuVal.....ThrIleIleGlyLeuAsnSerIst 55
580 TCCCTCTCCCTCCCTCCACAAAAAATAGTTTGTGTCTTAAATTTTGT 531
    |||:|||||:|||||:|||||
55 eunIsthrPro.....MetyrPhePheLeuPheAsnLeuSer 67
530 TTCTCTCTCTCTCTTATTCCTCCCTTAAATAATATATCTCTCTC 481
    |||:|||||:|||||:|||||
68 PheIleAspLeuCysTyrSerSerValPheThrProLysMetLeuMetAs,84
    |||:|||||:|||||:|||||
480 TTTTCATATTTTGAATCTCTCTCTCTCTAATAAATAAATAACACG 431
    |||:|||||:|||||:|||||
84 nPhe.....IleSerGluLysAsnIleIleSerPheI 95
430 GCATGTGTTT.....TTTGTATTTCTGTTTCTGTTATATCATCTT 391
    |||:|||||:|||||:|||||
95 ysgLysMetThrGlnLeuPhePheLeuPheLeuValIleSer 111
390 CCTGTGTGATATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 355
    |||:|||||:|||||:|||||
112 Glucys...TyrValLeuThrSerMetAlaTyrAspArgValAlaIleCy 127
354 .....CCCCCTCATATTAATAATCTCTTTTATTAATTTGGGGGTGAT 309
    |||:|||||:|||||:|||||
127 sAsnProLeuLeuTyrHisIleAlaMetSerProThrValCysSerSerI 144
308 AATTTTGGGGGAAAAAATCATAGTCCCCCATCGGGGATAGCTCCNCAC 259
    |||:|||||:|||||:|||||
144 eumetPheGlySer..... 148
258 ANGTCACGTTGTGCTATACCTGTTCACAAACANTGTGACACACAGT 209
    |||:|||||:|||||:|||||
149 .....TyrLeuMetAlaPheSerGlyAlaMetAlaHisThrGlyCy 162
208 TGCTTCATGTTG.....GTTGACACAGTCATCG 180
    |||:|||||:|||||:|||||
162 smetLeuArgLeuThrPheCysAspAlaAsnThrIleAspHisIstTyrPheC 179
179 TGAACAGCATGCCACACGTTCCACGACCTTGCTTCT 142
    |||:|||||:|||||:|||||
179 yAspIleLeuProLeuLeuGlnLeuSerCysThrSer 191

```

seq_name: /stds/gcgdata/geneseq/geneseqp/AA1990.DAT:AA05434

seq_documentation_block:

ID AAR05434 standard; protein; 555 AA.

alignment_scores: Length: 134
 Quality: 82.50
 Ratio: 1.650
 Gaps: 6
 Percent Similarity: 37.313 Percent Identity: 27.612

alignment_block:

US-09-823-101-3/rev x AAB51160 ..

Align seg 1/1 to: AAB51160 from: 1 to: 555

```

372 TGTCCCGCTCTCTCCCGCCCGCTCATATAAATATCTCTGTATTATA 323
    |||||
22 CysProvalGluAlaGluAlaProser..... 30
322 TTGGTGGGGTGATTAATTTTGGGGGAAAAACATCATGTCCCGCCCGCG 273
31 ..... SerThrProThrValProThrSerc 39
272 GGATAGCTCCNCAC.....ANGTCACGTTGTGCTATA..... 241
    |||
39 ysgluAlaLysGluGlyGlyCysIleAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTCAACCAACANTGTGACACACACAG 212
    |||||
56 ArgAspIleLeuSerAspGlyLeuCysGluAsnLysProGlyLysThrCy 72
211 TTGTGCTCTCATGTTGGTTGACACAGTCATCGTGAACGACGACACAG 162
    |||
72 scys..... 73
161 TTCACGCACTTGCTGTCTGTGCATGATATGATGAGATTAATTAAG 112
    |||
74 .....ArgMetCysGlnTyValIleGlnSerArgValGlu..AlaAlaGlu 88
111 ATAGCGTAGACACGATGATGATGATGATGATGATGATGATGATGATG 62
    |||||
88 YTYrPheArgThrPheTyArgValAlaLysArgPheAsnPheGlnGluProGlyL 105
61 AAGCACACCTGTGTGACAGCATAGGTCACACAGAGAACCCAGTGGACA 12
    |||
105 ystyValleuAlaLysArgGly.....ThrLysGlyGlyAspTrpSerc 118
seq_name: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.AAM36104

```

seq_documentation_block:

ID AAM36104 standard; Protein: 597 AA.

AC AAM36104;

DT 25-MAR-1998 (first entry)

DE Luciferase/epidermal growth factor receptor fusion protein.

KW Sea firefly; Vargula sp.; luciferase; label; mouse; murine; epidermal growth factor receptor; luminescent enzyme; fusion protein.

XX Chimeric - Vargula sp.

OS Chimeric - Mus sp.

FH Key Location/Qualifiers

FT Domain 1..555

FT Region /note= "Vargula luciferase"

FT Domain /note= "murine IgM heavy chain CH4 domain"

FT Domain /note= "EGFR transmembrane domain"

FT Domain /note= "EGFR cytoplasmic domain"

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

FT Domain

PF 25-AUG-1995; 95JP-0216911.

XX 25-AUG-1995; 95JP-0216911.

XX (TORA) TORAY IND INC.

XX WPI; 1997-492889/46.

XX A method of labelling cells - comprising a luminescent protein fused

XX to a trans-membrane receptor

XX Example 5; Page 6-7; 9pp; Japanese.

XX This sequence represents a fusion protein of the invention comprising the

XX sea firefly (Vargula sp.) luciferase linked to an epidermal growth factor

XX receptor (EGFR) transmembrane domain, via a murine immunoglobulin (Ig)

XX constant heavy chain domain 4 (CH4) linker. The fusion protein can be

XX used to detectably label cells.

XX Sequence 597 AA;

SQ

alignment_scores: Length: 134
 Quality: 82.50
 Ratio: 1.650
 Gaps: 6
 Percent Similarity: 37.313 Percent Identity: 27.612

alignment_block:

US-09-823-101-3/rev x AAM36104 ..

```

372 TGTCCCGCTCTCTCCCGCCCGCTCATATAAATATCTGTATTATA 323
    |||||
22 CysProvalGluAlaGluAlaProser..... 30
322 TTGGTGGGGTGATTAATTTTGGGGGAAAAACATCATGTCCCGCCCGCG 273
31 ..... SerThrProThrValProThrSerc 39
272 GGATAGCTCCNCAC.....ANGTCACGTTGTGCTATA..... 241
    |||
39 ysgluAlaLysGluGlyGlyCysIleAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTCAACCAACANTGTGACACACACAG 212
    |||||
56 ArgAspIleLeuSerAspGlyLeuCysGluAsnLysProGlyLysThrCy 72
211 TTGTGCTCTCATGTTGGTTGACACAGTCATCGTGAACGACGACACAG 162
    |||
72 scys..... 73
161 TTCACGCACTTGCTGTCTGTGCATGATATGATGAGATTAATTAAG 112
    |||
74 .....ArgMetCysGlnTyValIleGlnSerArgValGlu..AlaAlaGlu 88
111 ATAGCGTAGACACGATGATGATGATGATGATGATGATGATGATGATG 62
    |||||
88 YTYrPheArgThrPheTyArgValAlaLysArgPheAsnPheGlnGluProGlyL 105
61 AAGCACACCTGTGTGACAGCATAGGTCACACAGAGAACCCAGTGGACA 12
    |||
105 ystyValleuAlaLysArgGly.....ThrLysGlyGlyAspTrpSerc 118
seq_name: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.AAM61910

```

seq_documentation_block:

ID AAM61910 standard; Protein: 473 AA.

AC AAM61910;

DT 17-SEP-1998 (first entry)

```

DE D. melanogaster glutamate transporter.(1) sequence.
XX
KW Glutamate transporter; nervous system; research reagent;
KW arthropod; Drosophila melanogaster.
OS Drosophila melanogaster.
XX JPI0099083-A.
XX PD 21-APR-1998.
XX PF 27-SEP-1996; 96JP-0290990.
XX PR 27-SEP-1996; 96JP-0290990.
XX PA (SUNR ) SUNTORY LTD.
DR WP1: 1998-289876/26.
DR N-PSDB; AAVJ37697.
XX
PT Glutamate transporter obtained from arthropod - useful as reagent in
re research on nervous system
XX
PS Claim 7; Pages 5-7; 15pp: Japanese.
XX
CC This represents a Drosophila melanogaster glutamate transporter. The
CC specification provides glutamate transporter genes from an arthropod,
CC as shown in AAVJ37697 and AAVJ37698 and the corresponding protein
CC sequences are shown in AAW61910 and AAW61911. A glutamate transporter
CC gene encoding a glutamate transporter containing an amino acid sequence
CC in which at least 1 amino acid is added, deleted or replaced in the amino
CC acid sequence encoded by the above genes can be used in a method for
CC screening a glutamate transporter inhibitor where the glutamate
CC transporter gene is expressed in a culture cell and its glutamate
CC transporter activity is measured. The transporter can be used as a
CC reagent for the research on the nervous system.
XX
SO Sequence 473 AA:

alignment_scores:
      Quality:   81.50          Length:    130
      Ratio:     1.216         Gaps:       5
Percent Similarity: 51.538      Percent Identity: 28.462

Alignment Block:
US-09-823-101-3/rev x AAW61910 ..

Align seg 1/1 to: AAW61910 from: 1 to: 473

653 CCCCACCAATAATTTCTCTCGGCGTCCTCATAGCTGCATAATAC 604
|||||:::|||||||:::|||||||:::|||||||:::
17 ProProleuservalcysleuleuValasPserSerlaserSerly 33
603 TAAATATTCTCTCG.....TGCTGTCCCTCCTCCCC 569
|||:::|||||
33 stlealeualaSerelyarargluarGerCys.....ThrtYrP 48
568 CCCCACCAAAAAAATAGTTTTGTTCTTAATTTTCTTTTCTACTCTCT 519
||| ||| ||| ||| ||| ::| |||:::|||||
48 roserProalalySllepheluArghetleuScysleulleval... 63
518 CCTTAATCCCCCCCCCTTTAAAATAATTAATCTCTCTTCTTTTCAATTTG 469
||| ::| :|||:::|||||::|::|
64 .....ProleuValSerSerleThrSerAla.....IleG1 75

468 TAATCTCTCTCTCCCTAATAAAAAAAAAAACAACGCCGATGTTT 419
:::|||| |::|:::|||||:::|||||
75 yglyleuAspleuSermetSerlylleAlaThrArgAlaIlethr 92

418 TGATTTCTGTTTTCGTTTATCCATCTCTGTTGAATTTCTGTGTC 369
||||| ::|::| ::|::| ::|::|
92 yTrYrPheValThnThrIleSerAlaValIleleugIyIleCysleuVal 108

```

```

seq_documentation_block:
ID AAG71446 standard; Protein: 247 AA.
AC AAG71446:
AT
CG 31-JUL-2001 (first entry)
CT
DE Human olfactory receptor polypeptide, SEQ ID NO: 1127.
EN
KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation.
OS Homo sapiens.
PN MO200127158-A2.
PD 19-APR-2001.
PE 06-OCT-2000; 2000WO-US27582.
PF 08-OCT-1999; 99US-0158615.
PR 24-FEB-2000; 2000US-0184809.
PS
XX (DIGI-) DIGISCENTS.
XX (YEDA) YEDA RES & DEV CO LTD.
XX
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI; 2001-290713/30.
XX
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
XX Claim 11; Page 650-651; 1857pp; English.
XX
XX The present sequence is an olfactory receptor which is encoded by
XX one of a number of novel polynucleotides. The polynucleotides can be
XX used in screening for olfactory agonists and antagonists. The methods
XX allow for the determination of primary scents and the identification
XX of the odour receptors used to detect these primary scents. The methods
XX also enable determination of secondary scents and the identification of
XX combinations of odour receptors that are involved in detecting such
XX secondary scents. This enables the construction of a scent representation
XX (also called a scent fingerprint or scent profile), which may be used to
XX re-create and edit scents. Libraries of olfactory receptors are useful
XX for determining the interaction pattern of a composition with the
XX receptors, and can be used for determining differences in the olfactory
XX faculties of different individuals.
XX
XX Sequence 247 AA:

alignment_scores:
Quality: 80.00 Length: 169
Ratio: 0.976 Gaps: 8
Percent Similarity: 48.521 Percent Identity: 23.077

alignment_block:
US-09-823-101-3/rev x AAG71446 ..

Align seg 1/1 to: AAG71446 from: 1 to: 247

```

[illegible][illegible]

OM of: US-09-823-101-3 to: Issued_Patents_AA: * out_format : pfs

Date: Jan 17, 2002 4:01 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frimmet_n2p.model -DEV=xlp
-Q=/cgn2_1/USPSPool/US09823101/runtat.17012002_155137_22014/app-query.fasta_1.2239
-DB=Issued_Patents_AA -OPMT=fastan -SUFFIX=n2p.ra1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsum2 -TRANS=human40.cdi -LIST=45 -DOCCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09823101 -CGN1_1_59 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-3
Query length: 698
Database: Issued_Patents_AA: *
Database sequences: 212252
Database length: 22503292
Search time (sec): 66.610000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-811-897A-56				81.00	139.40	0.1958
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-058-389A-4				73.50	133.65	1.119
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-665-259-25				72.00	117.20	2.65
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-762-500-25				72.00	117.20	2.65
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-762-500-75				72.00	117.09	2.66
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-07-814-964-11				71.50	124.61	2.39
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-258-442-11				71.50	124.61	2.39
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-338-809-6				71.50	124.61	2.39
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-726-214-8				71.00	119.67	3.08
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-868-373-4				69.50	113.68	5.07
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-976-255-11				68.50	138.77	3.06
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-905-223-417				68.50	123.18	4.80
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-07-672-483-4				68.00	124.25	5.119
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-118-270-71				68.00	121.99	5.54
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:PCT-US93-08528-71				68.00	121.99	5.54
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-09-076-651-2				68.00	121.99	5.54
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-891-640-3				67.50	120.78	6.40
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-693-457-2				67.50	120.78	6.40
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-693-457-4				67.50	120.78	6.40
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-265-731-2				67.50	120.78	6.40
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-265-731-4				67.50	120.78	6.40
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-619-554-4				67.50	102.97	10.70
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-718-388-7				67.50	95.58	13.24
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-718-388-9				66.50	117.22	8.82
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-888-988-2				66.50	117.22	8.82
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-09-362-633-2				65.50	118.56	10.55
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-07-476-640-22				65.50	107.40	14.57
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-09-058-489-23				65.50	106.00	15.17
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-09-058-489-24				65.50	105.17	15.54
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-619-554-2				65.50	101.83	17.11
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-619-554-2				65.00	123.31	10.26
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-506-553C-8				65.00	123.49	10.51
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-506-553C-23				65.00	120.02	11.28
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-506-553C-26				65.00	119.18	11.56
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-765-856-2				65.00	119.18	11.56
/cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-935-009A-2				65.00	119.11	11.58
/cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-08-935-009A-4				65.00	119.11	11.58

/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-788-539A-2 - 65.00 117.83 12.02 343
/cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-07-609-510B-16 + 65.00 113.51 13.62 527
/cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-811-949-39 + 65.00 113.51 13.62 527
/cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:PCT-US91-01025A-2 + 65.00 113.51 13.62 5
/cgn2_6/ptodata/2/1aa/Backfiles1.pep:5185259-8 + 65.00 113.51 13.62 527 i Pa

seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-811-897A-56

seq_documentation_block:
Sequence 56, Application US/08811897A
Patent No. 5858787

GENERAL INFORMATION:

APPLICANT: ONDA, Haruo
APPLICANT: OHTAKI, Tetsuya
APPLICANT: MASUDA, Yasushi
APPLICANT: KITADA, Chieko
APPLICANT: ISHIBASHI, Yoshihiro
APPLICANT: HOSOTA, Masaki
APPLICANT: OGI, Kazuhito
APPLICANT: MIYAMOTO, Yasunori
APPLICANT: HABATA, Yugo
APPLICANT: SHIMAMOTO, No. 5858787io
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR PREPARING
TITLE OF INVENTION: SAID PROTEIN, AND USE THEREOF
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,897A
FILING DATE: 05-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/202,986
FILING DATE: February 25, 1994
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, David S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 44168-DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440

INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-811-897A-56

alignment_scores:

Quality: 81.00 Length: 253
Ratio: 0.757 Gaps: 11
Percent Similarity: 42.292 Percent Identity: 20.553

alignment_block:

US-09-823-101-3/rev x US-08-811-897A-56
Align seg 1/1 to: US-08-811-897A-56 from: 1 to: 1324

690 TACTGAGATTATTTTGTTCATTAATAGCGGCCCCCACTAAATA 641
|||||:|||||:|:|

```

1 TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
2 TITLE OF INVENTION: USE
3 NUMBER OF SEQUENCES: 22
4 CORRESPONDENCE ADDRESS:
5 ADDRESSEE: David A. Jackson, Esq.
6 STREET: 411 Hackensack Ave, Continental Plaza, 4th
7 STREET: 411 Hackensack
8 CITY: Hackensack
9 STATE: New Jersey
10 COUNTRY: USA
11 ZIP: 07601
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/058, 389A
19 FILING DATE: April 9, 1998
20 CLASSIFICATION: 800
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Jackson Esq., David A.
23 REGISTRATION/DOCKET NUMBER: 26,742
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 201-487-5800
26 TELEFAX: 201-343-1684
27 INFORMATION FOR SEQ. ID NO: 4:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 456 amino acids
30 TYPE: amino acid
31 STRANDEDNESS: single
32 TOPOLOGY: linear
33 MOLECULE TYPE: protein
34 DESCRIPTION: hENT1
35 HYPOTHEICAL: NO
36 FRAGMENT TYPE:
37 ORIGINAL SOURCE:
38 ORGANISM: Homo sapiens
39 US-09-058-389A-4
40
41 alignment_scores:
42 Quality: 73.50 Length: 235
43 Ratio: 0.758 Gaps: 14
44 Percent Similarity: 41.277 Percent Identity: 22.553
45
46 alignment_block:
47 US-09-823-101-3/rev x US-09-058-389A-4 ..
48
49 Align seg 1/1 to: US-09-058-389A-4 from: 1 to: 456
50
51 648 ACTAATAATTTCTCTCGGCGCTTCCTCATAGTCATTAATACATAA 599
52 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 36 ThrGlnrYrPhehrrAsnArgLeuAspMet.....SerGlnAs 48
54
55 598 TTATCTCTCTGTC.....GTCT 582
56 |||:|||||:|||||:|||||:|||||:|||||:|||||:
57 48 hvalSerleuvalThralaGluLeuSerLysaspAlaGlnAlaSerAla 65
58
59 581 CTCCTCTCTCCGCC...CCACACAAAAAATACCTTTTGTCTTAAATT 535
60 |||:|||||:|||||:|||||:|||||:|||||:|||||:
61 65 IaprolAlaIaProleupProGluArgGnsrleuSerAlaIePheAsn 81
62
63 534 TTGTTTCTACTCTCTCTTATTCGCCCTTTAAATAATTAATTACTGT 485
64 |||:|||||:|||||:|||||:|||||:|||||:|||||:
65 82 AsnvalMetThrLeuGlySalMetLeuProLeu.....Le 93
66
67 484 CTCTGTTTGCATATTGTGAATCTCTCTGCTAATAAAAAAATAATAC 435
68 |||:|||||:|||||:|||||:|||||:|||||:|||||:
69 93 uIeupheThrIyrlieu.....AsuSerP 101
70
71 434 ACGGCGCAT..... 427

```

```

101 heLeuNIhSGIaNgIleProclnSerValaArgIleLeuNgIySerLeuVal 117
111
426 ..SGTGTTTTGTATTTGCTGGTTTGGTTATATGCATCTTCTGTGGATA 380
118 AlaIleLeuValaPheLeuIleThrAlaIle..... 128
379 TTTCCGTGTCCTCCGTCTCTCCCTCCCTCC...CCCTCATATAAATATCTC 333
129 ...LeuValIyValGlnLeuAspAlaLeuProPhePheValIleThrM 144
332 TTGTAT.....ATTAAATGTGGGGT.....GGATAA 307
144 eTlleLyIleValleuIleLeuIleAsnSerPheGlyAlaIleuGInglySer 160
306 ATTTTGGGGGAAAAACTCATGTCCTCCCATCG..... 274
161 LeuPheGlyLeuAlaGlyLeuLeuProIaSerTyrThrIaProIleu 177
273 .....GGATATAGCTCCNCACAGTATACAGTGTGCTATACTGT 234
177 tSerIyGInglyLeuAlaGlyPhePheAlaSerValAlaMetIleCysa 194
233 CAACCAATGTGGACAC..... 217
194 IaIleAlaSerGlySerCluLeuSerGluSerAlaPheGlyTyrPheIle 210
216 ACAAGTTGTGCTCATGTTGTTGGTGGACCACTATGCTGAACGACATGCC 167
211 ThrAlaGSAIaValIleIle.....LeuThrIleIleCysTyrIle 224
166 ACACGTCGACGCACTGCTTGTCTGTGCATGATGATGACGTGAGAAT 117
224 uGlyLeuProArg.....LeuGluPheT 232
116 ATAGG 112
232 yTaIg 233

seq_name: /cgn2_6/ptdata/2/1aa/6A.COMB.pep:US-08-665-259-25
seq_documentation block:
: Sequence 25, Application US/08665259
: Patent No. 6028173
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burn, Timothy C.
: APPLICANT: Connors, Timothy D.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: APPLICANT: Klinger, Katherine W.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/665,259
: FILING DATE: 17-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315

```

```

1 REFERENCE/DOCKET NUMBER: IG5-9.1
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: (508) 872-8400
4 TELEFAX: (508) 872-5415
5 INFORMATION FOR SEQ ID NO: 25:
6 SEQUENCE CHARACTERISTICS:
7 LENGTH: 1684 amino acids
8 TYPE: amino acid
9 TOPOLOGY: linear
10 MOLECULE TYPE: protein
11 US-08-665-259-25

```

```

alignment_scores:
    Quality:      72.00          Length:      116
                Ratio:      1.263        Gaps:       7
Percent Similarity: 49.138     Percent Identity: 28.448

```

```

alignment_block:
US-09-823-101-3/rev x US-08-665-259-25 ..
Align seg 1/1 to: US-08-665-259-25 from: 1 to: 1684

```

```

690 TACTTGACATTATTTTGTCTTCATTAATAGCGCCCCCACAATAA 641
   ::::| | | | | | | | | | | | | | | | | | | | | |
289 PheLeuLeuPhePheLeuPheLeu..... 296
640 TTTCCTGGGCGTCCTCCCTCATAGCGCATATACTATAAT.....T 597
   ||||| | | | | | | | | | | | | | | | | | | | | |
297 ..... LeuIleAlaIleSerPheMetThrLeuLeuP 307
596 ATTCTCTGCTGCTGCTCCCTCC.....TCCCSCCC 565
   :: :| | | | | | | | | | | | | | | | | | | | |
307 hecysVallysvallysProasnValalavalLeuSeratgSeraSpro 323
564 ACCAAAAAATAGTTTGTGCTTAAATTTTGTCTCTCTCTCTCT 515
   ::::| | | | | | | | | | | | | | | | | | | | | |
324 SerLeuValleuAlaPheLeuLeuCysPheAlaIleSerThrIleSer... 339
514 ATTCGCCCCCTTTAAAATAATTACTCTCTCTCTTTTCATATTGCAAT 465
   || | | | | | | | | | | | | | | | | | | | |
340 .PheSerPheMet.....ValSerThrPhePheSer..... 349
464 CTCCTCTGCCTAATAAAAAAATAATACCACGGCGATGTTTTTGTGA 415
   || | | | | | | | | | | | | | | | | | | | |
350 ..... LysAlaAsnMetAlaAlaAlaIlePheglyly 359
414 TTTCGTTTTCGTTTATTCATCTCTCTGTTGATATTCTCTGTCGCC 367
   || | | | | | | | | | | | | | | | | | | | |
369 PheLeuTyrrPhePheMetThyrlleProtyr...PheheAlaIlaPro 374

```

```

seq_name: /cgn2_6/plodata/2/1aa/6A.COMB.pep:US-08-762-500-25
seq_documentation_block:
Sequence 25, Application US/08762500
Patient No. 6030806
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dockowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHOD OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
City: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,500
: FILING DATE: 09-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/665,259
: FILING DATE: 17-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10469
: FILING DATE: 17-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: 1G5-9.3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 872-8400
: TELEFAX: (508) 872-5415
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1684 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-762-500-25

```

```

alignment_scores:
: Quality: 72.00 Length: 116
: Ratio: 1.263 Gaps: 7
: Percent Similarity: 49.138 Percent Identity: 28.448

```

```

alignment_block:
: US-09-823-101-3/rev x US-08-762-500-25 ..
: Align seg 1/1 to: US-08-762-500-25 from: 1 to: 1684

```

```

690 TACTTGAGATATTTGTTGTTCTTCATTAAGCGCCCCCACTAATAA 641
: : : : : : : : : : : : : : : : : : : : : : : : : : :
289 PheLeuLeuPhePheLeuLeu..... 296
640 TTCTCTCGGCGTCCCTCATAGTCGATATATACTAAT.....T 597
297 .....LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 307
596 ATTCTCTCGTGTCTCCCTCC.....TCCCCCCC 565
: : : : : : : : : : : : : : : : : : : : : : : : : : :
307 hecysValIysValIysProAsnValAlaValLeuSerArgSerAspPro 323
564 ACCAAAAAATAGTTTGTCTTAAATTTGTTTCTACTCTCTCCTT 515
: : : : : : : : : : : : : : : : : : : : : : : : : : :
324 SerLeuValLeuAlaPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 339
514 ATTCCCCCCCCCTTAAATAATTAATCTCTCTTTCATATTTGTAA 465
: : : : : : : : : : : : : : : : : : : : : : : : : : :
340 .PheSerPheMet.....ValSerThrPhePheSer..... 349
464 CTCCTCTCGCTAATATAAAAAAATAACACCGCATGTGTTTGTGA 415
: : : : : : : : : : : : : : : : : : : : : : : : : : :
350 .....LysAlaAsnMetAlaAlaAlaPheGlyGly 359
414 TTCTGTTTTCGTTTATTCATCTTCCTGTGATTAATTCCTGTGTC 367
: : : : : : : : : : : : : : : : : : : : : : : : : : :
360 PheLeuTyPhePheThrTyIleProTy...PhePheValAlaPro 374

```

```

seq_name: /cgn2_6/plodata/2/iaa/6A_COMB.pep:US-08-762-500-75
seq_documentation_block:
: Sequence 75, Application US/08762500

```

```

: Patent No. 6030806
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burr, Timothy C. D.
: APPLICANT: Connors, Timothy R.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: APPLICANT: Klinger, Katherine W.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
: NUMBER OF SEQUENCES: 83
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,500
: FILING DATE: 09-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/665,259
: FILING DATE: 17-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10469
: FILING DATE: 17-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: 1G5-9.3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 872-8400
: TELEFAX: (508) 872-5415
: INFORMATION FOR SEQ ID NO: 75:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1704 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-762-500-75

```

```

alignment_scores:
: Quality: 72.00 Length: 116
: Ratio: 1.263 Gaps: 7
: Percent Similarity: 49.138 Percent Identity: 28.448

```

```

alignment_block:
: US-09-823-101-3/rev x US-08-762-500-75 ..
: Align seg 1/1 to: US-08-762-500-75 from: 1 to: 1704

```

```

690 TACTTGAGATATTTGTTGTTCTTCATTAAGCGCCCCCACTAATAA 641
: : : : : : : : : : : : : : : : : : : : : : : : : : :
309 PheLeuLeuPhePheLeuLeu..... 316
640 TTCTCTCGGCGTCCCTCATAGTCGATATATACTAAT.....T 597
317 .....LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 327
596 ATTCTCTCGTGTCTCCCTCC.....TCCCCCCC 565
: : : : : : : : : : : : : : : : : : : : : : : : : : :
327 hecysValIysValIysProAsnValAlaValLeuSerArgSerAspPro 343
564 ACCAAAAAATAGTTTGTCTTAAATTTGTTTCTACTCTCTCCTT 515
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

344 SerLeuValLeuAlaPheLeuLeuCySPheAlaIleSerThrIleSer. 359
514 ATTCGCCCCCTTTAAATAATTAATCTCTCTCTTTTCATATTTGTAAT 465
||| ::::: |||||
360 .PheSerPheMet.....ValSerThrPhePheSer..... 369
464 CTCCTCTGCTTAATAAAAAAATAACACGCGCATGTGTTTTTGTA 415
370 .....LysAlaAsnMetAlaAlaAlaPheGlyGly 379
414 TTTCGTCTTTCGTTATCATCTCTCTGTTGATATTTCCGTCTCC 367
|||||::: ||| ::::: ||| ::::: |||
380 PheLeuTyrrPhePheThrTyrrIleProTyr...PhePheValAlaPro 394

seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:us-07-814-964-11

seq_documentation_block:
; Sequence 11, Application US/07814964
; Patent No. 5359047
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pili, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellelt, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P. C.
; STREET: 2 Millfilia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; IMMEDIATE SOURCE:
; CLONE: Drosophila SSRP (predicted)
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 458..507
; OTHER INFORMATION: /label= Acidic
; FEATURE:

```

```

; NAME/KEY: Domain
; LOCATION: 518..547
; OTHER INFORMATION: /label= Basic I
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 547..620
; OTHER INFORMATION: /label= HMG-box
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 632..649
; OTHER INFORMATION: /label= Basic II
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 657..723
; OTHER INFORMATION: /label= Mixed Charge
; US-07-814-964-11

alignment_scores:
Quality: 71.50 Length: 129
Ratio: 1.212 Gaps: 5
Percent Similarity: 45.736 Percent Identity: 23.256

alignment_block:
US-09-823-101-3 x US-07-814-964-11 ..
Align seg 1/1 to: US-07-814-964-11 from: 1 to: 723

349 GGGGGGGGGGAGAGACGCGGACACAGCAATATCAACAAAGAAATGCA 398
|||||::: |||||::: |||||::: |||||::: |||||:::
511 GlyIleGlyAspSerAspGlyAlaLysLysLysLysLysSergI 527
399 TAAAGAAACAGCAATACAAACAAACAC..... 426
::: |||||::: |||||::: |||||::: |||||::: |||||:::
527 uLysLysGluLysLysLysLysLysLysLysLysLysLysLysLysL 544
427 .....ATCGCGTGTATT 441
|||||:::
544 ySProSerLysLysLysLysAspSergLysProLysArgAlaThrThr 560
442 TTTTITTTT.....TTAGCGAGAGAGATTACAAATATGAAAGAGA 485
|||||::: |||||::: |||||::: |||||::: |||||:::
561 AlaPheMetLeuTrpLeuAsnAspThrArgLysSerLysLysLysLys 577
486 G.....AGTATTATTATTAAAGCGGCGGGAATTAAGAG 520
577 nProGlyIleLysValThrGluIleAlaLysLysGlyGluMetTrpL 594
548 G.....ACAAAACTATTTTTGGTGGCGGGGA.. 543
521 AGAGTAAACAAACAAATTAAG..... 543
::: |||||::: |||||::: |||||::: |||||::: |||||:::
594 ySGluLeuLysAspLysSerLysTrpGluAspAlaAlaLysAspLys 610
544 .....AACAAAACTATTTTTGGTGGCGGGGA.. 573
611 GluArgTyrrHisAspGluMetArgAsnTyrrLysProGluAlaGlyLys 627
574 .....CGAGGAGACACAGCAGAGAGA 594
627 pSerAspAsnGluLysGlyLysSerSerLysLys 639

seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:us-08-258-442-11

seq_documentation_block:
; Sequence 11, Application US/08258442
; Patent No. 5670621
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pili, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellelt, Patti
; APPLICANT: Essigmann, John M.

```

```

349 GGGGGGGGAGAGACGGCGACAACGAATATCAACGAAGAANGA 398
      |||||||::: |||||:::~::~ ~::|||:::
511 GtlylglYlslYspseraspGlYalAlYslYslYslYslYsl 527
      ::::::::::|| ||||||
527 uylslYsgUlYslYsgUlYslYshSlYslYsgUlYgLuArthrYsl 544
      ::::::::::::::::::::::::::::ATGCCGTGCTATT 441
427 .....
544 yspRoserlYslYslYsAspSerGlYslYsProlYsArgAlAthrThr 560
      ||:::::
442 TTTTTTTT.....TTAGCAGAGAGAGATTCAAAATARGAAAAGA 485
      ||::::: ||::::: |||::::: ||:::::
561 AlaphmeclEurpleuSnsapThrArgCluserlrlYslArgGuas 577
      ::::::: |||::::: |||::::: ||:::::
486 G.....AGTAATTATTATTAAAAAGGGGGGGAATAAGAG 520
      :
577 nProglYlelYslValThrclutleAlAlYslYsgUlgUmetrrPl 594
      ::::::: |||||||
521 AGAGTGAAAAACAAAATTAAAG..... 543
      ::::: ||::::: |||
594 YsgUleuYslAspLysSerlYsrTrpGUlsPaLaAlAlYslAspLys 610
      ::::: ||::::: |||
544 .....AACAAACTATTTTTGGTGGGGGGGA.. 573
611 GluArGTyrHisApclumetaArgstYrLYslPrGUlaAgLYslAs 627
574 .....GAGCGAGACACAGCAGAGAGA 594
      |||||||::: ||:::::
627 pseraspasnGLuLyglYlYslYsserlYslYs 639

seq_name: /cgn2_6/p/odata/2/iaa/5A.COMB.pep:US-08-328-809-6

seq_documentation_block:
Sequence 6, Application US/08328809
Patent No. 5705334
GENERAL INFORMATION:
APPLICANT: Lippard, Stephen J.
APPLICANT: Essigmann, John M.
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pil, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kelleck, Patli
TITLE OF INVENTION: Uses For DNA Structure-Specific
TITLE OF INVENTION: Recognition Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Patent Administrator, Testa, Hurwitz & Thibault
STREET: 53 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328, 809
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fenton, Gilliam M.
REGISTRATION NUMBER: 36, 508
REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-248-7100
TELEFAX: 617-248-7100
```


OTHER INFORMATION: /label= Mixed Charge
PCT-US92-11107-11

Align seg 1/1 to: PCT-US92-11107-11 from: 1 to: 723

seq_documentation_block:

Sequence 8, Application US/08726214
Patent No. 6107076
GENERAL INFORMATION:
APPLICANT: Tang, Wei-Jen
APPLICANT: Gilman, Alfred G.
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435

```
alignment_scores:
  Quality: 71.00      Length: 175
  Ratio: 0.887      Gaps: 11
Percent Similarity: 45.714      Percent Identity: 25.143
```

alignment_block:
US-09-823-101-3/rev x US-08-726-214-8

Align seg 1/1 to: US-08-726-214-8 from: 1 to: 1064

[illegible]


```

CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 417:
SEQUENCE CHARACTERISTICS:
LENGTH: 92 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig-peptide
LOCATION: -41...-1
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 4.7
OTHER INFORMATION: seq LFSCFPLSHKFG/KK
US-08-905-223-417

alignment_scores:
Quality: 68.50 Length: 91
Ratio: 1.712 Gaps: 3
Percent Similarity: 43.956 Percent Identity: 32.967

alignment_block:
US-09-823-101-3/rev x US-08-905-223-417 ..

Align seg 1/1 to: US-08-905-223-417 from: 1 to: 92

539 AATTTTGTTCACCTCTCTATATCCCCCTTTAAAAATAATT 430
||| |||:||||| ||| |||||
3 AsnAlaCysIleValLeuProPthProPro..... 14
489 ACTCTCTCTTCATATTGTATCTCTGCTCAATAAAAAAA 440
||| :||||| |||||
15 ....SerLeuGlnProSerAlaSerLeuAlaProAsn..... 26
439 TAACACGGCAGTGTCTTTTGTATTCTGTTTCTTATCCAGCTTC 390
|||: ||| ||||| |||
27 .....ArgPheLeuPheSerCysPheCysPheLeuSerHisLysPhe 40
389 CTGTGTGATTTTCCTGTCGCCGCTCTCTCC....CCCCATAT 343
::: ||||| |||:
41 GILysIysValIleTyrPheAsnTyrLeuSerGluLeuHisGluHisLe 57
342 AAATATCTCTTGTATTATTAATGCGGGGAGATTAATTTTGGGGAAA 293
:||||| :|||: |||
57 uuySTyRAspGlnLeuValIleProGluValLeuArgTyrAspGluL 74
292 AACTCATGTCCTCCCATCGGGA 270
||||: ||| |||: |||

```

```

74 yslEugInserLeuHisGluGly 81
seq_name: /sgn2_6/Plodata/2/1aa/5A.COMB.pep:US-07-672-483-4
seq_documentation_block:
Sequence 4, Application US/07672483
Patent No. 5359029
GENERAL INFORMATION:
APPLICANT: LACROIX, Martial
TITLE OF INVENTION: PEPTIDES AND ANALOGUES AND MIXTURES
THEREOF FOR DETECTING ANTIBODIES TO HTLV-I AND HTLV-II
TITLE OF INVENTION: VIRUSES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/672,483
FILING DATE: 19910302
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554,258
FILING DATE: 18-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: HALEY Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: 1AF8 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0674
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-672-483-4

alignment_scores:
Quality: 68.50 Length: 110
Ratio: 1.457 Gaps: 6
Percent Similarity: 42.727 Percent Identity: 26.364

alignment_block:
US-09-823-101-3/rev x US-07-672-483-4 ..

Align seg 1/1 to: US-07-672-483-4 from: 1 to: 433

529 TTCTACTCTCTCTTATTCCTCCCTTTAAAAATAAT.....TA 489
||| :||| :|||: |||: |||
49 PheLeuLysLeuAlaLeuLysThrProIleTyrPheAsnProIleAspTy 65
488 CTCTCTCTTTCATATTGTATCTCTGCTCAATAAAAAAAAT 439
||||| ||| :||| |||
65 rSerLeuLeu.....AlaSerLeuIleProLysGlyT 76
438 AACACGGCAGTGTCTTTTGTATTCTGTTTCGT..... 401
||| ||| |||: |||: |||
76 yrProGluArgValValGluIleIleAsnIleLeuValLysAsnGluVal 92
400 .....TATCATCTCTCTGTGATATTTC 375

```

us-09-823-101-3.n2p.rpr

p1r2:S5015
 p1r2:T2004
 p1r2:I4583
 p1r2:T0537
 p1r2:A4154

```

seq. document
Cypripidia
N;Alternat
C;Species
C;date: 09
C;Accessio
R;Thompson
Proc. Natl
A;Title: C
A;Referenc
A;Accessio
A;Status:

```

A;Residues
A;Cross-re
C;Keywords
alignment_

[illegible]

seq_name	pir2:A33723
pir2:S50150	- 71.00 124.02 15.08 427 i gastric CCK-A receptor - rat
pir2:T20045	- 71.00 122.80 15.14 498 i hypothalamic protein C49A1.5 - bovin
pir2:I45832	+ 71.00 119.91 15.26 715 i scinderin - bovine
pir2:T05374	+ 71.00 119.43 15.28 760 i hypothetical protein F16G20 - Drosophila
pir2:A41542	- 71.00 116.74 15.40 1064 i adenylate cyclase EC 4.6.1.1 - human

```

seq_documentation_block
Cyclindina-luciferin 2-monooxygenase (EC 1.13.12.6) precursor - ostracod (Vargula hilgendorfii)
N:Alternate names: crustacean luciferase
C:Species: Vargula hilgendorfii (sea firefly)
C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 26-May-2000
C:Accession: A33723
R:Thompson, E.M.; Nagata, S.; Tsuji, F.I.
Proc. Natl. Acad. Sci. U.S.A. 86, 6567-6571, 1989
A:Title: Cloning and expression of cDNA for the luciferase from the marine ostracod Vargula hilgendorfii
A:Reference number: A33723; MUID:89367290
A:Accession: A33723
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-555 <THO>
A:Cross-references: GB:M25666; NID:g162546; PID:g162547
C:Keywords: luminescence; monooxygenase; oxidoreductase

```

alignment_scores:	
Quality:	90.50
Ratio:	1.775
Percent Similarity:	38.060
	Length: 134
	Gaps: 6
	Percent Identity: 28.358

```
alignment_block:
US-09-823-101-3/rev X A33723
..
```

Align seg 1/1 to: A33723 from: 1 to: 555 .

372 TGTCCCGCTCTCTCTCCCGCCCCCTCATATAAATACTCTGTATTAA 323
|||||||::: |||||||

322 TTTGGTGGGGGATAAATTTTGGGGGAAAACTCATGTCCCCCCATCGG 273

```

31 .....SerThProThValProThSerc 39
      :::  |||||  |||

```

[illegible]

240CTCTGTTCAACCAANTGTGACACACAAG 212

56 ArgAspIleLeuSerAspGlyLeuCysGluAsnLysProGlyGlyThrCys 72

72	SCVS	73
71		
70		
69		
68		
67		
66		
65		
64		
63		
62		
61		
60		
59		
58		
57		
56		
55		
54		
53		
52		
51		
50		
49		
48		
47		
46		
45		
44		
43		
42		
41		
40		
39		
38		
37		
36		
35		
34		
33		
32		
31		
30		
29		
28		
27		
26		
25		
24		
23		
22		
21		
20		
19		
18		
17		
16		
15		
14		
13		
12		
11		
10		
9		
8		
7		
6		
5		
4		
3		
2		
1		
0		

161 TTCCACCACCTTGCTGCTGCTCATGCATATGACTGTGAGACATTATAGG 112
|||||.....|||

113 ArgMetCysGlnTyrValIleIleIleCysArgValGluAlaAlaG1 88
74

```

      ||| |||
88 ytyrPheargTrPheuryrGlyLysArgPheasnPheGlnGluProGlyL 105

```

61 AAAGCACACCTGGTGTGACAGGGCATATGGGTCACACAGCAGGAACCCAGTGGACA 12
|||::: ||| ||| ::|||:::
|||::: ||| ||| ::|||:::

```
sed name: pfr2:T47488
```

seq_documentation_block: *unorthodox* protein 60 - Arabidopsis thaliana

C;species: *Arabidopsis thaliana* (mouse-ear cress)


```

305 TTTTGGGGGAAAACTCATGTCCCGCATGGGAGTACGTCNCACANG 256
263 Ser.....:
255 TCACGTTGTGCTACTCTGTTCAACCAANTGTGACACACAGTTGGC 206
267 lPheIleuLeuLysLysIleValThr.PhePheIleHisThrIle... 282
205 TCTCATGTTGTTGACACAGTCATGTCATGATGATGATGATGATGAT 160
283 .....ArgHisArgGluLeu.MetLysHisValHisL 293
159 .....CCACGACATTGCTGTCTGTCATGATGATGATGATGATGAT 118
293 yValLeuProArgLysLeuThrLysLeuIleAsnPhaSplacInser 309
117 TAT 115
310 His 310

```

seq_name: p1r2:H86373

```

seq.documentatation.block:
protein T33E23.16 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86373
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86373
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-653 <STO>
A:Cross-references: GB:AE005172; MID:9369395; PIDN:AAF87143.1; GSPDB:GN00141
C:Genetics:
A:Gene: T33E23.16
A:Map position: 1

```

```

alignment_scores:
Quality: 82.00 Length: 180
Ratio: 0.965 Gaps: 11
Percent Similarity: 47.222 Percent Identity: 23.889

```

alignment_block:
US-09-823-101-3 x H86373 ..

Align seg 1/1 to: H86373 from: 1 to: 653

```

11 CTGTCCACTGGG.....TTCTGCTGTGACCCATGCGCTGCACAG 54
|||||
511 LeuSerThrGlyAspLysMetLeuGlnTrpAsnArgLysSerProTh 527
55 T...GTGCTTGCCTGGAT.....CTTCTTGA 80
|||||
527 rcysValLeuLysAsnAsnIleGlnThrArgAsnHisLeuPhePhe 544
81 GTCTGTGTCATGACATCTGTCCTAGCGTATCCATTAATTTCACAGTC 130
|||
544 erCysCysTyrThrAla..... 549
131 ATATGATGACAGACAGACAGAGTGGTGGACGTGTGCATCATTAC 180
|||||

```

```

550 ...GluIleTrpGluAsnLeuAlaLysAsnIleTyrLysAlaLysPhe 565
181 GATGACGTGTGCATCCACATGAGAGCAACATTGTGTCACANTGC. 229
:|||||
565 rThrsnTrpSerThrIleLeuThrSer.....ValSerThrTrp 580
220 .....TTTGACAGAGTATAGCACACAGTACNTGTGNGAGCTATC 271
|||
580 rGAsnArgThrGluSerPheLeuAlaArgTyr..... 590
272 CCCGATGGGGGACATGAGTGTTCCTCCCAAAATTTATCCACCCACCA 321
591 .....IlePheGlnAlaThrIleHis..... 597
322 ATTATTAACAAGATATTTATATGAGGGGGGGAGAGAGAGGAGAC 371
|||
598 .....ThrIleTrpHisGluArgAsnGlyArgThrGly 610
372 ACAGGAAATATCACACAGAGAGA.....TGATTAACGAAAC... 409
610 lvarGSerAsnSerAlaThrHisLeuIleTrpTrpLeuAspLysGlnMet 626
410 AGAATACAAAAACACATGCCCGGTATTTTATTTATAGCGAG 459
|||||
627 ArgAsnGlnIleSerThrIleAla.....AlaSer 636
460 AGGAGATTACAAATATGAAAGAGAGAGATATTTATTTT 499
|||||
636 rGlyAspHisArgTyrAspLysProLeuGlnLeuTrpPhe 649

```

seq_name: p1r2:H70108

```

seq.documentatation.block:
hypothetical protein B80072 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: H70108
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: H70108
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-773 <KLE>
A:Cross-references: GB:AE001120; GB:AE000783; MID:92687951; PIDN:AAC66467.1; PID:9268
A:Experimental source: strain B31

```

```

alignment_scores:
Quality: 82.00 Length: 135
Ratio: 1.155 Gaps: 7
Percent Similarity: 52.593 Percent Identity: 28.889

```

alignment_block:
US-09-823-101-3/rev x H70108 ..

Align seg 1/1 to: H70108 from: 1 to: 773

```

680 TATTTTGTCTTCATTAATAGTCGGCCCGCATTAATATTCTCTCGG 631
|||||
425 TyrPheAlaIleIleLeuSerLeuSerProSerLeuIlePheSerPhe.. 440
630 CGTCTCCCTTCATAGCTGCATTAATACATAATTAATTCCTGCTGTC 581
|||||
441 .....LeuAsnAsnThrLysPheMetThrAspThr.... 450
580 TCCCTCTCTCCCGCCCGACCAAAATAATAGTTTCTCTTAATTTTGT 531
|||||
451 .....LeuLeu.LeuIlePheIle 456

```

```

530 TTTCTACTCTCTCTATTCCTTATCCCTTAAATAAATATCTCTTC 481
      :      :      :      :      :      :      :      :
456 eSerValAlaIleTyrIleProIleMetLeuLysnTyrIlyscysL 473
480 TTTTCATATTTGTAAT.....CTCTCTCTGCTTAAT 450
      ||      :      :      :      :      :      :
473 eUSeRpheValTyrAlaAlaLeuMetIleSerIleIleProLeu 489
449 AAAAAAATAAAC.....ACGCCAGTGTG 424
      :      :      :      :      :      :      :
490 AspleuGluIleThrLeuLysIleIleSerLeuLeuPheThrPheLeu 506
423 TTTTTCATTTTCGTTTTCGTTTATCCATCTCTCTGATATTTCT 374
      |||||      :      :      :      :      :      :
506 lPhePheTyrPhe...PheThrLeuSerAlaPheLeuAlaasn..... 519
373 GTGTCGGCTCTCTCTCCCTCATATATAATATCTCTTTATTA 324
      ||      :      :      :      :      :      :
520 .....ProIleLeuIleSerPheMetPheLeuVal 529
323 ATT 321
      ::
530 Leu 530
seq_name: p1r2:T02001

```

```

seq_documentation_block:
hypothetical protein T15B16.10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C:Accession: T02001
R:Stonking, T.; Smith, R.
submitted to the EMBL Data Library, November 1998
A:Description: The sequence of A. thaliana T15B16.
A:Reference number: Z14488
A:Accession: T02001
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-652 <STO>
A:Cross-references: EMBL:AF104919; NID:g3859590; PID:g3859608
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 135/1
A>Note: T15B16.10

```

```

alignment_scores:
  Quality: 81.00      Length: 209
  Ratio: 0.931      Gaps: 17
Percent Similarity: 41.627      Percent Identity: 25.837

```

```

alignment_block:
US-09-823-101-3/rev x T02001 ..

```

```

Align seg 1/1 to: T02001 from: 1 to: 652

```

```

584 TGTCCTCCCTCTCCSCCCSCACCA..... 561
      |||||      :      :      :      :      :
105 CysAlaLysTyrProProGluValIleAsnIleSerLysThrIleHis 121
560 .....AAAAATAGTTTGTCTTAAATTTTGTGTTTC 527
      ||      :      :      :      :      :
121 shiLysLeuAsnLeuLeuLysGluIleGlnPheAspTyrCysVal. 137
526 TACTCTCTCTATTCCTCCSCCTTAAATAAATATCTCTCT..... 483
      :      :      :      :      :      :
138 .....TpnIsrProGluValAsnshIsthrLeuGluValAsn 150
482 ...TCPTTTCATATTTGTAATCTCTCT.....CTGCC 454
      |||||      :      :      :      :      :
151 HisSerLysIleSerLeuIleProLeuLysLeuHisThrGluIlePur 167

```

```

453 TAAATAAAAAATAACACAGCCATGTGT..... 423
      |||||      :      :      :      :      :
167 oAsnTyrSerAspArgLysCysArgLeuCysAlaLysGluIleGluValG 184
422 ..TTTTGATATTCCTGTTTTCGTTTATCCATCTCTCTGTAATTTCC 375
      :      :      :      :      :      :
184 lYeuPheTyrIleHisCys...SerLeuCysasnPheThrIleAspMetSer 199
374 TGTCCTCCCTCTCTCTCCSCCTCATATAAATATCTC..... 333
      |||||      :      :      :      :      :
200 CysVal.....LeuasnProProGln....ArgTyrIleThrPysnIle 212
332 .....TTGTAAATTAATGTTGGGCGATGAATTTT 302
      |||||      :      :      :      :      :
212 ulysAlaHisAspHisGlnIleuThrLeuProSerLeuArgSerPheI 229
301 TGGGGGAAAAATCATCATGCTCCCGGAGATAGCTCCACAGATCAC 252
      ||      :      :      :      :      :
229 eu.Cys...serAlaCysGlyLeuAsnGlyAspArgSerProTyrIleC 244
251 GTTG...TGCTATATCTCTGTTCAACCAANTGTGGACACACAGTTGCT 205
      :      :      :      :      :      :
244 sValGlnCysAspPheMetIle.....HisGlnValCysL 256
204 CTCATGTTGGTGCACACATCATGTCGTAACCTGCATGCCACAC..... 163
      |||||      :      :      :      :      :
256 euHisLeu.....ProArg.ValIleAsnIleAsnArgHisAspHisAr 270
162 .GTTCACGACGACT.....T 150
      |||||      :      :      :      :      :
270 gValSerArgThrSerValLeuGlyValAlaAsnSerValCysGlyValC 287
149 GCCTTGCTGCATCGATATGACT 127
287 ysArgGlnLysValAspTyrPthr 294
seq_name: p1r2:T37929

```

```

seq_documentation_block:
probable major facilitator protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T37929; T38038
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21763
A:Accession: T38038
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-166 <MUR>
A:Cross-references: EMBL:Z98598; PIDN:CA811245.1; GSPDB:GN00066; SPDB:SPAC1B3.19
A:Experimental source: strain 972h-; cosmid c1B3
C:Genetics:
A:Gene: SPDB:SPAC1B3.19
A:Map position: 1

```

```

alignment_scores:
  Quality: 79.50      Length: 88
  Ratio: 1.622      Gaps: 3
Percent Similarity: 55.682      Percent Identity: 25.000

```

```

alignment_block:
US-09-823-101-3/rev x T37929 ..

```

Align seg 1/1 to: T37929 from: 1 to: 408

```

588 TTATTCCTGCTGCTGCTCCCTCCCTCCACCAAAAAATAGTTT 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 Leu1leAla1yGln1Ala1yVal1Leu1sP1roAs1Gln1Val1Leu1Ser1Th 122
548 TTGTTCTTTAATTTTGTTCCTCTCTCTCTATTCCTCCCTCC 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
122 rCys1erPro1Leu1Trp1Le1Ser1Val1lle1Ty1rLeu1Leu1s1nPro1Leu1Th1P 139
505 .....CCTTAAATAATAATACCTCTCT 482
133 he1euPro1G1y1leAla1Cys1Ser1Ala1sP1me1l1le1Leu1s1nPh1e1Th1r1Th 155
481 CTTTTCATATTTGTAATCTCCCTCTG.....CTAATAAAAAAAA 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 Leu1Met1Th1lle1Ty1rPh1eAla1Ser1Cys1G1Ser1Ty1rAla1lle..... 169
440 ATATACACGCGCATGTTTGTATTTCTGTTTCGTTATTCATCTT 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 .....Ty1rAla1Cys1Met1Ala1leu1Th1rVal1Ph1e1le1As1nPro1As1n 184
390 CCTTGTGATATT 377
184 la1eu1leu1Ph 188

```

seq_name: p1r2:T44522

seq_documentation_block:
 hypothetical protein 4S (Imported) - Shigella sonnei
 C:Species: Shigella sonnei
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T44522
 R:Chida, T.; Okamura, N.; Yoshida, Y.; Ohtani, K.; Arakawa, E.; Watanabe, H.
 submitted to the EMBL Data Library, May 1999
 A:Description: Complete DNA sequence of the O-antigen (rtb) gene cluster in Plesiomonas
 A:Reference number: 222787
 A:Accession: T44522
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-426 <CH1>
 A:Cross-references: EMBL:AB028135; PIDN:BA85073.1
 A:Experimental source: Isolate HM383

alignment_scores:

Quality:	77.50	Length:	190
Ratio:	0.891	Gaps:	7
Percent Similarity:	45.789	Percent Identity:	22.632

alignment_block:

US-09-823-101-3/rev x T44522 ..

Align seg 1/1 to: T44522 from: 1 to: 426

```

677 TTTTCTTTCATTAAGTCGCGCCGCCACATAATTTCTCGGCGT 628
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 Ph1e1Met1Ph1e1Ty1r1Ser1Ty1Val1Ser1sP1Ser1le1Ser1leu1Th1leu1Va 109
627 CTCCTCTCATAGTCGATTAATTAATTAATTAATTTCTCTGCTGCTGCC 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 lTy1r1le1Ser1Ser1Ala1Ty1rPh1eAla1Cys1lle1Ser1Ser1G1y1rPh1e1Ser 125
577 CTCGCCCCCCCCCCACCAAAA.....AATAGTTT 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 Leu1eu1Gln1Ala1Val1G1y1Ar1Ph1eAr1Gys1Glu1Leu1Ty1r1le1As1n1le1Ty 142
548 TTGTTCTTTAATTTTGT.....TTTCTACTCTCTCTTAATTC 511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 rSer1Th1r1lle1le1Ty1r1lle1G1y1Cys1As1n1leu1Ph1e1leu1Ser1le1le1G 159
510 CCCCCCTTTAAATAATTAATTAATTAATTAATTTGTAATCTCC 461

```

```

159 luoPro.....LeuTyTyTySer..... 164
460 TCCTGCTAATAAAAAAATAAACACGCGCATGTGTTTTGTATTTTC 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 .....Ala1lle1Se 167
410 TGTTTTCGTTATTCATCTTCCTTGTGATATTTCTGTCGCTGCTCT 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 r1le1Ph1e1lle1Ty1r1Ser1le1Ser1leu1Val1Ph1e1Ser1Ser1H1s1Ty1Cys1A 184
360 CTCGCCCCCCCTCATATAAA..... 339
184 sn1Val1Pro1Cys1Ph1e1H1s1le1Ty1sAr1G1Pro1Ser1le1leu1Val1Ty1r1Ly1s1As1P 200
338 .....TATCTCTGTTATTAAT 322
201 Ph1e1leu1sP1Ala1Th1rPro1Ph1eAla1lle1leu1Val1leu1leu1As1nVal1le 217
321 TGTGGGGGTGATAAATTTTGGGGGAAACATGTCGCCCATCGGG 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 uSer1le1leAs1P1leu1Ph1e1lle1leu1Ty1G1uTy1rPh1e1Ser1Ty1rAs1n.... 232
271 GATFACCTGCNCACANGTCACGTTGTGCTACTCTGTTCAACACANTGTG 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 .....SerVal1Ala1lle1Ty1rGln1Val1Val1Th1r1G1Val 243
221 GACACACAAGTTTGCTC 204
244 Asn1Th1G1ly1leu1le1r1le 249

```

seq_name: p1r2:T21502

seq_documentation_block:
 hypothetical protein F28D9.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21502
 R:Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19431
 A:Accession: T21502
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-384 <M1>
 A:Cross-references: EMBL:Z81518; PIDN:CAB04215.1; GSPDB:GN00019; CESP:F28D9.2
 A:Experimental source: Clone F28D9
 C:Genetics:
 A:Gene: CESP:F28D9.2
 A:Map position: 1
 A:Introns: 56/3; 97/3; 127/3; 227/2; 293/1

alignment_scores:

Quality:	77.00	Length:	204
Ratio:	0.837	Gaps:	9
Percent Similarity:	45.098	Percent Identity:	22.549

alignment_block:

US-09-823-101-3/rev x T21502 ..

Align seg 1/1 to: T21502 from: 1 to: 384

```

524 CTCCTCCCTTATTCGCCGCC...TTTAAATAATAATTAATCTCT 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
7 l1e1Cys1Pro1Ty1rG1y1Pro1Pro1As1Py1Ty1r1Met1Ty1s1Th1r1leu1Gln1leu1 23
482 .....TCCTTCATATTTGTAATC 464
23 ec1y1leu1Ph1e1Ser1leu1Pro1le1As1nPh1e1leu1Cys1Ph1e1Ty1rPh1e1Val1Trp 40
463 TCCTCCTGCTAATAAAAAAATAAACACGCGCATGTGTTTTGTAT 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

40 heYrThrProLysGluSerLysPhe.....ArgTYrCysLeuAlaTyr 54
413 TTGCGTTTGCGTTATGCCATCTCTGTTGATATTTCCGTGTC...CC 367
||| ||| :||| |||||:|||| ||| ||
55 PheGlnPheIleAlaPheLeuValGluValAspMetSerLeuValCyspr 71
||| ||| :||| |||||:|||| ||| ||
366 GTCTCTCTCCSCCCCCCATATATAATCTCTGTTATTAATGCG 317
|||||:|||||:|||||:|||||:|||||
71 OGlyPhe.....TyrLeuPheProLeuMetGlyG 81
||| ||| :||| |||||:|||| ||| ||
316 GCGTGCATTAATTTTGGGGAAAACATGATGCCCCCATCGGGGATAG 267
|| :|||:|||||:|||||:|||||
81 LyrTrpAsnLeuAlaGluThrAsnArgLeuPheSerGlyHisGln..... 95
266 CTCNCACAGTGCACGTGTGCTACTCTGTTCAAAACANTGACAC 217
|||||:|||||:|||||:|||||:|||||
96 .....ThrValValPheGlyTyrPheLysPheSerPheGluLe 108
216 ACAAGTTGCTCTCATATGTTGTTGACACGTCATCGTAACATGCA... 171
:||||| |||||:|||||:|||||
108 uProSerLeuLeuLeuLeuSerPheIleTyrArgHisAsnAlaIleAsnpr 125
||| ||| :||| |||||:|||| ||| ||
171 ..... 171
125 heAsnProLysPheLysIleArgArgSerIleGlnTyrLeuValIleAla 141
170 ...TGGCAGACGTTCCAGCAGCATGCGTGT..... 144
|||||:|||||:|||||:|||||:|||||
142 ThrCysHisThrPheProPheValThrAlaIleCysLeuLeuLysSerG 158
||| ||| :||| |||||:|||| ||| ||
143 CTGTGCATCATATGACTGTTGAGAAATTATAGATACGTAAG..... 102
|||||:|||||:|||||:|||||:|||||
158 uLeuSerHisValGlnGlnValArgIleLeuGluLysValArgArgTyrL 175
101 .....ACCAGATGCATAGCAGAGACTCACAAGAATCCAGGCCAAA 60
|||||:|||||:|||||:|||||:|||||
175 ysrLeuGlnThrArgGluThrAlaAsnTyrThrGlnAsnTyrProAsn 191
59 GCACACCTGGTG 48
|||||:|||||:|||||:|||||:|||||
192 LeuHisIleLeu 195
seq_name: p1r2:T44511

```

```

seq_documentation_block:
hypothetical protein 3p [imported] - Plesiomonas shigelloides
C:Species: Plesiomonas shigelloides
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44511
R:Chida, T.; Okamura, N.; Yoshida, Y.; Ohtani, K.; Arakawa, E.; Watanabe, H.
submitted to the EMBL Data Library, April 1999
A:Description: Complete DNA sequence of the O-antigen (rfb) gene cluster in Plesiomonas
A:Reference number: 222786
A:Accession: T44511
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-421 <CH1>
A:Cross-references: EMBL:AB025970; PIDN:BAA85008.1
A:Experimental source: ATCC 14029

```

```

alignment_scores:
Quality: 75.50 Length: 190
Ratio: 0.868 Gaps: 7
Percent Similarity: 45.789 Percent Identity: 22.632

```

alignment_block:

US-09-823-101-3/rev x T44511 ..

Align seg 1/1 to: T44511 from: 1 to: 421

```

677 TTTTGTTCCTCATATAGTCGCCCCCACAATAATTCCTCGCGGT 628
||| |||:|||||:|||||:|||||:|||||

```

```

93 PheMetPheTyrSerTyrValSerAspSerIleSerLeuThrLeuLeuVal 109
627 CTCCTCTCTCATCTGCATATACATAATATATCTCTCTGCTGCTCC 578
||| ||| :||| |||||:|||| ||| ||
109 lTyrIleSer.SerAlaTyrPheValCysIleSerSerGlyArgPheSer 125
||| ||| :||| |||||:|||| ||| ||
577 CTCCTCCCCCCCCACCAAAA.....AATAGTTT 549
|||||:|||||:|||||:|||||:|||||
126 LeuLeuGlnAlaValAlaGlyArgPheArgCysGluLeuTyrIleAsnIleTyr 142
548 TTGTTCTTAAATTTTGT.....TTCTACTCTCTCTTATTC 511
||| ||| :||| |||||:|||| ||| ||
142 rSerThrIleIleTyrIleCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAAAATTAATTAATCTCTCTCTTTCATATTTGTAATCTCC 461
||| ||| :||| |||||:|||| ||| ||
159 lupro.....LeuTyrTyrSer..... 164
460 TCTGCTTAATAAAAAAATAAACAACGCGCATGTTTGTGTAATTC 411
||| ||| :||| |||||:|||| ||| ||
165 .....AlaIleSer 167
410 TGTTCCTTTATTCATCTCTCTGTTGATATTTCTGTCGCTCT 361
||| ||| :||| |||||:|||| ||| ||
167 rIlePheIleTyrSerIleSerLeuLeuValPheSerSerHisLysCysA 184
360 CTCCTCCCCCCCCCT.CATATAAA..... 339
||| ||| :||| |||||:|||| ||| ||
184 snValProCysPheHisIleLysArgProSerLeuLeuValTyrLysAsp 200
338 .....TACTCTGTTAATTAAT 322
||| ||| :||| |||||:|||| ||| ||
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValIle 217
321 TGTGCGGTGATTAATTTTGGGGAAAACATCATGTCCTCCCATCGGG 272
||| ||| :||| |||||:|||| ||| ||
217 uSerSerIleAspLeuPheIleLeuLysGluTyrPheSerTyrAsn.... 232
271 GATAGCTCCACACAGTCAGTGTGCTATACCTGTCTCAACACANTGTC 222
||| ||| :||| |||||:|||| ||| ||
233 .....SerValAlaIleTyrGlnValValThrArgVal 243
221 GACACACAGTTGTGCTC 204
||| ||| :||| |||||:|||| ||| ||
244 AsnThrClyLeuMetIle 249
seq_name: p1r2:A83133

```

```

seq_documentation_block:
probable MFS transporter PA4096 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83133
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; M01D:20437337
A:Accession: A83133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE004826; GB:AE004091; NID:9950296; PIDN:AA607483.1; GSPDB:GN
A:Experimental source: strain PA01
A:Gene: PA4096

```

```

alignment_scores:
Quality: 75.00 Length: 150
Ratio: 1.136 Gaps: 7
Percent Similarity: 44.000 Percent Identity: 25.333

```

alignment_block:

US-09-823-101-3/rev x A83133 ..

Align seg 1/1 to: A83133 from: 1 to: 419

```
662 ATAGTGGCCCCCACAATAATTTCTCGGCGTCCCTCTCATAGCT 613
      :::::|||||
154 VALLIEGIPROPROLEUVALALALEUALAGUALALEUGLYTRPAR 170
      :::::|||||
612 GCATAATCTAATATATCTCTGCTGCTC..... 579
      ||| |||
170 GHISALAEHELLEVALSERLEUPROGLYLEUUEUILEALTRPLEUI 187
      ||| |||
578 .....CCTCCCTCCCCCACCMAAAAVAGTTTGG 546
      ||| |||
187 IETPRAGHISVALARGPROASPROPRORALAGLARG..... 200
      ||| |||
545 TTCCTTAATTTTGTCTACTCTCTCTTAATCCCTTTAAAAA 496
      ::|||
201 .....PROLAPROARGLALAGLUALALYSGI 210
      ::|||
495 TAAATTAATCTCTCTTTCATATTTGTAATCTCCTCGCTATAATAA 446
      ||| |||
210 YLSARGLEUALALEUULYSERARGASNILLEU..... 223
      ||| |||
445 AAAAAATACACGCGCATGTGTTTGTATTTCTGTTTCGTTATCC 396
      ||| |||
224 .....CysThrIleuIleSerCysValPheValIthr 233
      ||| |||
395 ATCTTCCTTTGATATTTCTGTGCTCCGCT..... 363
      ||| |||
234 TRPHEILEVALLEULIESERPHETHRPHLEUVALASNALAAR 250
      ||| |||
362 ....CTCCTCCCCCTCATATATAATATCTCTGTTATTAATGGTG 317
      ||| |||
250 GGLYHESERPROALATHRMETGLYARGLEUMETSERCYSLEU.GLYVAL 266
      ||| |||
316 CGGTGCAATAATTTTGGGGGAAAAACATCATGTCCCTCCATCGGGAT 269
      ||| |||
267 ALATRPVAL...VALTRPGLY...PHEALAYALPROALALIESERASP 280
```

seq_name: p1r2:E71551

seq_documentation_block:

probable monooxygenase - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999

C:Accession: E71551

R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra

A:Reference number: A71570; MUID:99000809

A:Accession: E71551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-507 <ARN>

A:Cross-references: GB:AE001288; GB:AE001273; NID:g3328545; PIDN:AMC67739.1; PID:g332854

C:Genetics:

A:Gene: mmpA

alignment_scores:

Quality:	75.00	Length:	159
Ratio:	1.056	Gaps:	8
Percent Similarity:	44.654	Percent Identity:	27.673

alignment_block:

US-09-823-101-3/rev x E71551 ..

Align seg 1/1 to: E71551 from: 1 to: 507

```
545 TTCTTTAATTTTGTGTTT..... 528
      ||| |||
215 PHEVALASNPHEVALPHETRYASNPROTYARGLYSERARGINLEU 231
      ||| |||
527 .....CTACTCTCTTAATTCCTCCCTCCCTTAAATAATATAC 488
      ::||| |||
231 rLEUALASnThrSerGlyProLeuSerSerLysPheLysAsnLysLeuL 248
      ||| |||
487 TCCTCTCTTTTCATATTTGTAATCTCCTCGCTAATAAAAAATA 438
      ||| |||
248 eUTYrThrTyGlyLeuAla.....LeuAlaGLuAspProLeuSerIle 262
      ||| |||
437 ACCACGGCCATGTGTTTTGTATTTCTGTTTCGTTATTCATCTCTCT 388
      ::|||
263 SerSerSerLeuLeuGlnTyPrProPheCysHisAspArgTyrIlePhe 279
      ||| |||
387 TGTGATATTTTCCTGTGTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 338
      ||| |||
279 uGLYSerIleAlaAsnAsnLeuSerPheSer..... 289
      ||| |||
337 ATCTCTGTATTAATTTGTTGGGGTG.....GATAAATTT 303
      ::|||
290 .....TyrlleuSerGlyAlaAsnSerAsnIleHisAspAlaPhe 302
      ||| |||
302 ..TTGGGGGAAAACTCATGTATCCCTCCATCGGGGATAGCTCCACANG 256
      ||| |||
303 AsnleuGlyTrpLysLeuLeu.ProvalIleLysLysAlaIleSerSerg 319
      ||| |||
255 TCACGTGTGCTATACCTGTTCMAACCANVTGACACACACAGTGTGC 206
      ||| |||
319 InLeuIleLeuSerLysGluLeuLys.....Thr 328
      ||| |||
205 TCTCATGTGTTGGTTGCACACGTCATGTCGTAATGTCATGCCACGTTCCAC 156
      ||| |||
329 SerHisVal...LeuProHisPheAsnGluValHisGlnLysAlaGln 344
      ||| |||
155 GCACCTGCTGTCTGTGCATCGCATAT 131
      ||| |||
344 LysLeuLeuPheSerAsnMetLys 352
```

seq_name: p1r2:C81704

seq_documentation_block:

monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain N19g)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: C81704

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39

A:Reference number: AB1500; MUID:20150255

A:Accession: C81704

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <JET>

A:Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39281.1; PID:g719

C:Genetics:

A:Gene: TC0425

alignment_scores:

Quality:	74.50	Length:	156
Ratio:	0.993	Gaps: <td>8</td>	8
Percent Similarity:	48.077	Percent Identity:	26.923

alignment_block:

US-09-823-101-3/rev x C81704 ..

Align seg 1/1 to: C81704 from: 1 to: 506

551 TTTTGTCTTTAAT.....TTTGTGTTT 529

```

|||||:|||||:
218 PheValPheTyrAsnProTyrArgGlySerArgGlnLeuTyrLeuAlaAs 234
528 TCTACTCTCTCCCTATTCCTCCCTTTAAATAATTAATCTCTCTT 419
234 ntrSerGlyProLeuSerProLysPheAsnLysLeuLeuTyrThr 251
478 TTCCATATTGTAAATCTCTCTGCTATATAAAAAATAAACACGCGC 429
251 YrGlyLeuAla.....LeuAlaGlnAsnProLeuSerLieserAlaSer 265
428 ATGTGTTTTTGTATTTCTGTTTCTTTATTCATCTCTCTGTATAT 379
266 PheMetGlnTyrProPheGlySHisAsnArgTyrLephLeuGlySerVa 282
378 TTCCCTGTGCTCCCTCTCTCTCCCTCCCTCATATATAATCTCTGT 329
282 LAlaAsnAsnLeuSerPheSer..... 289
328 TATTAATGCTGGGCTG.....GATAAATT...TTGGGG 297
290 ..TyrLeuSerGlyValAsnLeuAsnLiesGlnAlaPheAsnLeuGly 305
296 GAAAAACATCATGTCCCTCCATCGGGGATAGCTCCNCAACAGTCAGTTGT 247
306 TrpLysLeuLeu..ProValLileLysLysAlaAlaSerSerGlnLeuLeL 322
246 GCTATACTCTGTCTCAACACANTGTGGACACACAAGTTGTCTCTCATGTT 197
322 euserArgGlnLeuLys.....ThSerHisVal 331
196 GGTGACCATCATGCTGTAACATGCATGCCACAGCTTCACGACCTGCT 147
332 ...LeuProHisPheAsnGlnValHisGlnLysArgThrAlaLysLeuLe 347
146 TGCTGTCTCATCATGATAT 131
347 upHeserLysMetTyr 352
seq_name: p1r2:rt14191

```

```

seq_documentation_block:
  extensin homolog T28D5.60 - Arabidopsis thaliana
  C:Species: Arabidopsis thaliana (mouse-ear cress)
  C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
  R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
  submitted to the Protein Sequence Database, August 1999
  A:Reference number: 217931
  A:Accession: T14191
  A:Molecule type: DNA
  A:Residues: 1-350 <BEV>
  A:Cross-references: EMBL:AL109819
  A:Experimental source: cultivar Columbia; BAC clone T28D5
  C:Genetics:
  A:Gene: ATSP:T28D5.60
  A:Map position: 4
  C:Superfamily: hydroxyproline-rich glycoprotein

```

```

alignment_scores:
  Quality: 74.00      Length: 122
  Ratio: 1.370       Gaps: 7
  Percent Similarity: 44.262   Percent Identity: 30.328

```

alignment_block:

US-09-823-101-3/rev x T14191

Align seg 1/1 to: T14191 from: 1 to: 350

```

672 TTTCTTCATTAATAGTCGCCCCCCTAAATATTCTCTCTGCGGCTGCC 623
|||||:|||||:|||||:

```

```

241 PheIleTyrSerSerProProProProTyrValTyrLysSerValPr 257
622 TCTCATAGCGCATTAATACATATATTTCTGTCGTCGTCCTCCT 573
257 oArgLlePro.....PheIleTyrSer..SerPro.....ProPro 268
572 CCCCCCACCACAAAAAATAGTTTTTGTCTTAAATTTGTTTCTACT 523
269 ..ProProTyrValTyrAsnSerLarProArgLleProPheIle..TyrS 284
522 CTCTCCTTATTCCTCCCTTTAAATAATTAATCTCTCTTTTCAATA 473
284 eSerLeuProProProProTyrValTyrAsnSerLarProArgValPro 300
472 TTGTATATCTCTCTG.....CCTAATAAAAAATAATACAC 433
301 PheIleTyrSerSerProProProProProTyrValTyrAsnSerAlaPr 317
432 GCGCATGTGTTTTTGTATTTCTGTTTCTGTTATTCATCTCTGTTG 383
317 oArgLleProPheIleTyr..... 323
382 ATATTTCTGTGTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 333
324 .....SerSerProProProHisHisMetPheThr 333
332 TTGTTATTAAT 321
334 SerLeuPheLeu 337
seq_name: p1r2:B96724

```

```

seq_documentation_block:
  hypothetical protein F20P5.14 [imported] - Arabidopsis thaliana
  C:Species: Arabidopsis thaliana (mouse-ear cress)
  C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
  R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
  Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
  ansem, N.F.; Hughes, B.; Hultzer, L.
  Nature 408, 816-820, 2000
  A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
  C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitli, R.; Marzia
  Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
  A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
  ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
  A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
  A:Reference number: A86141; MUID:21016719
  A:Accession: B96724
  A>Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-760 <STO>
  A:Cross-references: GB:AE005173; NID:92194126; PIDN:AB61101.1; GSPDB:GN00141
  C:Genetics:
  A:Gene: F20P5.14
  A:Map position: 1

```

```

alignment_scores:
  Quality: 74.00      Length: 186
  Ratio: 1.057       Gaps: 10
  Percent Similarity: 37.634   Percent Identity: 24.194

```

alignment_block:

US-09-823-101-3/rev x B96724

Align seg 1/1 to: B96724 from: 1 to: 760

```

680 TATTTTGTCTTCTCATTAATAGCGGCC..... 651
|||||:|||||:|||||:
16 TyrPhe...PhePheIleValValLeuProPheGlnSerLeuSerGlnPh 31
650 .....CCACTAAATATTCTCTGCGGCTCTCCCTCTCATAGTCGCATA 608

```

```
31 eapSerProglInasnle.....G 38
    || |||||
607 ATACTAATAATTCCTCGTCGTCCTCCCTCCCCCACCACAAA 558
    ::||::: |||||
38 luthrPhePheProIleSerSerIeuSerProValProProPro..... 52
    |||||
557 AAATAGTTTTGTTCTTAAATTTTGTTCCTCTCTCCCTTA..... 513
    |||||
53 .....LeuLeuProProSerSerAs 59
    |||||
512 ....TCCCCCCCCCTTAAATAATTAATCTCTCTTTTCATATTTGTA 467
    |||||
59 nProSerProProSerAsnAsnSerSerSer..... 70
    |||||
466 ATCTCCTCCTCAATAATAAAAAAATAACACCGCCGATGTTT... 420
    ::|||
71 .....AspLysLysThrIleThrLysAlaValIleuIleThr 82
    |||||
419 .....TTGATTTCTGTTTCGTTT 400
    ::|||
83 AlaAlaSerThrIleuLeuValAlaGlyValPhePhePheCysIeuGlnAr 99
    ||::: |||
399 ATCCATCTTCCTT.....GTGATATT. 378
    ||::: |||
99 gCysIleIleAlaArgArgArgArgValGlyProValArgValG 116
    ::|||
377 ..TCCTGTCCCGTCTCTCCCGCCCT..... 348
    ::|||
116 lUAsrThrIeuProProTyProProProPrometThSerAlaAlaVal 132
    ::|||
347 .....CATATAA 340
    ::|||
133 ThrThrThrThrIleuAlaArgLugLypheThrArgPheGlyValLy 149
    ::|||
339 ATATCTCTTTTATTAAATGCGGGGCGATAAATTTGGGGGAAAAAC 290
    ||::: |||
149 sgLyIeuIleLeuAspGlnAsngLyIeuAspValIeuTyTrpArgLysL 166
    ::|||
289 TCATGTCC 282
    ||::: |||
166 eugInSer 168
```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-3 to: SwissProt_39:* out_format: pfs
Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame+n2p.model -DEV=xip
-O/cgn2.1/USPRO.spool/US09823101/runat.17012002.155138.22061/app_query.fasta_1.2239
-DB=SwissProt_39 -QEMT=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blcsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=pfis
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USPR=US09823101.ccgml_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-3
Query length: 698
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 66.470000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
SwissProt_39:LUCL_VARRH	-	90.50	157.93	0.0722	555	P17554 vargula hilgendorffii
SwissProt_39:YELJ_SCHRO	-	79.50	138.75	1.15	408	O13883 schizosaccharomyces F
SwissProt_39:OJ23_HUMAN	-	74.50	131.12	4.01	311	O76001 homo sapiens (human)
SwissProt_39:OB88_HUMAN	-	74.50	131.12	4.01	311	O15620 homo sapiens (human)
SwissProt_39:OEH3_HUMAN	-	73.50	129.01	5.18	316	O60404 homo sapiens (human)
SwissProt_39:EMT1_HUMAN	-	73.50	126.00	5.29	455	O09808 homo sapiens (human)
SwissProt_39:OATP_RAT	-	73.00	122.81	5.41	670	P46720 rattus norvegicus (rat)
SwissProt_39:VIPS_RAT	-	73.00	125.35	5.99	437	P35000 rattus norvegicus (rat)
SwissProt_39:OB82_HUMAN	-	72.50	127.17	6.66	311	O09286 homo sapiens (human)
SwissProt_39:VED1_RHPIV	-	72.00	131.82	7.26	157	P24834 rhesus papillomavirus
SwissProt_39:EBD1_ARATH	-	72.00	117.01	8.07	945	P42762 arabidopsis thaliana
SwissProt_39:AR33_HUMAN	-	72.00	112.14	8.35	1704	O09758 homo sapiens (human)
SwissProt_39:HYPA_HYPLI	-	71.50	126.80	8.48	256	P35587 hypodermia lineatus (G
SwissProt_39:YKR9_YEAST	-	71.50	118.75	8.98	679	P34237 saccharomyces cerevis
SwissProt_39:SSRP_DROME	-	71.00	118.23	9.02	723	O05344 drosophila melanogast
SwissProt_39:CKR_RABIT	-	71.00	121.59	9.92	427	O09772 cyctolagus cuniculus
SwissProt_39:CYA4_RAT	-	71.00	114.05	10.47	1064	P26770 rattus norvegicus (rat)
SwissProt_39:VEAR_HCVNA	-	70.50	128.21	10.67	170	P09694 human cytomagalovirus
SwissProt_39:O2D2_HUMAN	-	70.50	123.30	11.05	308	O08210 homo sapiens (human)
SwissProt_39:YCKA_BACSU	-	70.50	120.96	11.23	409	O08791 bacillus subtilis. hy
SwissProt_39:V1PR_RAT	-	70.00	119.02	12.83	459	P30083 rattus norvegicus (rat)
SwissProt_39:VAN9_SCHRO	-	70.00	116.07	13.11	656	O10074 schizosaccharomyces F
SwissProt_39:KAPB_YEAST	-	69.50	119.59	14.41	380	P06245 saccharomyces cerevis
SwissProt_39:ND4M_ASTPE	-	69.50	118.01	14.57	460	P11992 asterina pectinifera
SwissProt_39:NI42_SOYBN	-	69.50	112.56	15.15	890	P39870 glycine max (soybean)
SwissProt_39:NI41_ORISA	-	69.50	112.33	15.17	916	P16081 oryza sativa (rice)
SwissProt_39:CYB_SATSC	-	69.00	118.62	16.35	379	O035930 salmistr sciureus (comm
SwissProt_39:UNI_PEA	-	69.00	118.28	16.39	395	O048559 pisum sativum (garden pe
SwissProt_39:Y618_AQUAE	-	69.00	117.92	16.43	413	O067545 aquifex aeolicus. hyf
SwissProt_39:HOC_DROME	-	69.00	109.98	17.39	1080	O09286 drosophila melanogast
SwissProt_39:CDUG_BOVIN	-	68.50	120.94	18.12	254	O09286 bos taurus (bovine)
SwissProt_39:PCGA_HUMAN	-	68.50	116.71	18.68	424	O15235 homo sapiens (human)
SwissProt_39:GAG_HTLV2	-	68.50	116.54	18.70	433	P03346 human t-cell leukemia
SwissProt_39:ENV_SRV1	-	68.50	107.69	19.92	587	P04027 simian retrovirus str-1
SwissProt_39:SVY2_HUMAN	-	68.50	116.67	21.06	1264	P26640 homo sapiens (human)
SwissProt_39:MGK_HUMAN	-	68.00	115.47	21.06	378	P35410 homo sapiens (human)
SwissProt_39:VPS_MOUSE	-	68.00	105.40	22.82	437	P41588 mus musculus (mouse)
SwissProt_39:ROD_ODOSI	-	67.50	122.04	22.84	1481	P49468 odontella sinensis.
SwissProt_39:FRAP_CAEBL	-	67.50	115.07	24.00	175	P41855 caenorhabditis elegans
SwissProt_39:RRC_SALTY	-	67.50	115.07	24.00	407	P26479 salmoneilla typhimurium

SwissProt_39:ENV_HTLIF - 67.50 113.58 24.26 488 O03817 human t-cell leuken
SwissProt_39:GLS2_YEAST + 67.50 102.38 26.28 1895 P40989 saccharomyces cer
SwissProt_39:YMI6_NEPOL - 67.00 118.21 26.45 247 O01C94 neophoselmis oliva
SwissProt_39:Y917_AQUAE - 67.00 114.07 27.25 408 O067062 aquifex aeolicus.
SwissProt_39:YBRD_SCHRO - 67.00 111.65 27.72 547 O13747 schizosaccharomyce

seq_name: SwissProt_39:LUCL_VARRH

seq_documentation_block:

ID LUCL_VARRH STANDARD; PRT; 555 AA.
AC P17554;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE LUCIFERIN 2-MONOOXYGENASE PRECURSOR (EC 1.13.12.6) (CYPRIDINA-TYPE
DE LUCIFERASE).
OS Vargula hilgendorffii (Sea firefly) (Cypridina hilgendorffii).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Ostracoda; Mydocalopa;
OC Mydocalopidae; Cypridinoida; Cypridinidae; Vargula.
OX NCBI_Taxid=6674;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=89367290; PubMed=2771943;
RA Thompson E.M., Nagata S., Tsuji F.I.;
RT Cloning and expression of cDNA for the luciferase from the marine
RT ostracod Vargula hilgendorffii.
RL Proc. Natl. Acad. Sci. U.S.A. 86:6567-6571(1989).
CC -I- CATALYTIC ACTIVITY: LUCIFERIN + O(2) -> OXIDIZED LUCIFERIN + CO(2)
+ LIGHT.
CC -I- THE CYSTEINE RESIDUES PRESUMABLY EXIST IN INTRAMOLECULAR
CC DISULFIDE BRIDGES.
CC -I- MISCELLANEOUS: 60 KCAL/MOL ARE REQUIRED FOR THE BLUE EMISSION OF
CC LIGHT (460 NM) DUE TO THE OXIDATION OF LUCIFERIN VIA A DIOXETANONE
CC INTERMEDIATE, IN WHICH THE EXCITED STATE OXYLUCIFERIN BOUND TO
CC LUCIFERASE IS THE EMITTER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M25666; AAA30332.1; .
DR PIR: A33723; A33723.
DR InterPro: IPR001846; Vwd.
DR Pfam: PF00094; Vwd; 2.
DR SMART: SM00216; VWD; 2.
KW Oxidoreductase; Monooxygenase; Photoprotein; Luminescence;
KW Decarboxylase; Signal; Glycoprotein.
FT SIGNAL 1 11
FT CHAIN 12 555
FT MOD_RES 12 12
FT DOMAIN 39 82
FT CARBOHYD 186 186
FT CARBOHYD 408 408
FT SEQUENCE 555 AA; 61626 MW; 2FB150371A80A0DF CRC64;

alignment_scores:

Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6
Percent Similarity: 38.060 Percent Identity: 28.358

alignment_block:

US-09-823-101-3/rev x LUCL_VARRH
Align seg 1/1 to: LUCL_VARRH from: 1 to: 555

372 TGRCCGCTCTCTCTCCCCCCCCCTCATATAAATATCTGTATTAA 323
|||||
22 CysproValGluAlaGluProSer..... 30


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL022727: CA18783.1; -
DR EMBL: AJ302547: CAC20472.1; -
DR EMBL: AJ302548: CAC20473.1; -
DR EMBL: AJ302549: CAC20474.1; -
DR EMBL: AJ302550: CAC20475.1; -
DR EMBL: AJ302551: CAC20476.1; -
DR EMBL: AJ302552: CAC21440.1; -
DR EMBL: AJ302553: CAC21441.1; -
DR EMBL: AJ302554: CAC21442.1; -
DR EMBL: AJ302555: CAC21443.1; -
DR EMBL: AJ302556: CAC21444.1; -
DR EMBL: AJ302557: CAC20477.1; -
DR EMBL: AJ302558: CAC20478.1; -
DR InterPro: IPR000276: GPCR_Rhodopsn.
DR Pfam: PF00001: 7tm_1; 1.
DR PROSITE: PS00237: G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS00262: G_PROTEIN_RECEP_F1_2; 1.
KM G-protein coupled receptor; Transmembrane; Glycoprotein;
KM Multigene family; Olfaction; Polymorphism.
FT DOMAIN 1 28 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 29 52 1 (POTENTIAL).
FT DOMAIN 53 60 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 61 82 2 (POTENTIAL).
FT DOMAIN 83 103 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 104 123 3 (POTENTIAL).
FT DOMAIN 124 142 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 143 161 4 (POTENTIAL).
FT DOMAIN 162 198 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 199 222 5 (POTENTIAL).
FT DOMAIN 223 239 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 240 262 6 (POTENTIAL).
FT DOMAIN 263 275 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 276 295 7 (POTENTIAL).
FT DOMAIN 296 311 CYTOPLASMIC (POTENTIAL).
FT DISULFID 100 192 BY SIMILARITY.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 113 113 T->A (IN 6M1-3*02).
FT VARIANT 226 226 R->Q (IN 6M1-3*02).
FT VARIANT 228 228 V->I (IN 6M1-3*03).
FT VARIANT 261 261 I->M (IN 6M1-3*03 AND 6M1-3*04).
FT SEQUENCE 311 AA; 34950 MW; 84201145B9D5AA68 CRC64;

Alignment_scores:
Quality: 74.50 Length: 163
Ratio: 0.968 Gaps: 10
Percent Similarity: 47.239 Percent Identity: 23.926

Alignment_block:
US-09-823-101-3/rev x 02J3_HUMAN ..
Align seg 1/1 to: 02J3_HUMAN from: 1 to: 311

597 TATCTCTGTCGTCTGTCCTCTCTCC...CCCCCACCACAAAATAAGT 551
111 :::::::::::::::::::: 111 ::::
14 TyrPheIleIeuValGlyPheSerAsnTrpProHISLeuGluValValI 30
550 TTTCGTTTAAATTTTGTCTTCTACCTCTCCCTATTCCCCCCCTTT 501
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
30 ePheValIValIleuIlePheTyIleuMetHISLeuIleGlyAsnLeu 46

```

```

500 AAAAAATAATTAATCTCTCTTTTCAATTTGTAATCTCTCCGCTGA 451
::: ::::::::::::::::::::|
47 .....PheIleIleIleuSerTyIleu..... 54
450 TAAAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 410
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
55 .....AspSerHISLeuHISThrProMetTyIlePheIleuSerAs 68
409 GTTTCGTTTATTCATCTCTCTGTTGATATTCCTGTCGCCGTCGTC 360
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
68 nLeuSerPheIleuAsp.....LeuCyTYIleHIS 79
359 TCCCCCCCCCTCATATAATAATATCTCTGTTTAAATTAATTAATTA 310
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
79 eSerIlePro.....GlnLeuValAsnIleuTrpGly... 90
309 TAAATTTTGGGGGAAACATCATGTCCCCCATCGGGCATAGCTCCNCA 260
111 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
91 .....ProGluTyIleSerTyIle 98
259 CANGTCACGTTGCTATCTCTGTTCAACCAAGTGGACACACAGTT 210
::: ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
98 aGlyCysMetIleGlnLeuTyIleValIleuGlnGlyThrThrGlnC 115
209 GTGCTCTCATGTTGGT.....GACCACGTCATCTGGAACGATG 169
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
115 yValIleuLeuValIleuMetSerTyIleAspArgTyIleAlaValCysArg 131
168 CCA...CACGTT.....CCAGCAGCTGTC 148
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
132 ProIleuHISTyIleThrValIleuMetHISProArgPheCys 144

seq_name: SwissProt_39:08B8_HUMAN

seq_documentation_block:
ID 08B8_HUMAN STANDARD; PRT; 311 AA.
AC Q15620;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OLFACTORY RECEPTOR 8B8 (OLFACTORY RECEPTOR TPCR85).
GN OR8B8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 126-282 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97224452; PubMed=9119360;
RA Vanderhaeghen P., Schumann S., Vassart G., Parmentier M.;
RT "Specific repertoire of olfactory receptor genes in the male germ
cells of several mammalian species.";
RL Genomics 39:239-246(1997).
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X89675: CA61822.1; -
DR InterPro: IPR000276: GPCR_Rhodopsn.
DR Pfam: PF00001: 7tm_1; 1.
DR PROSITE: PS00237: G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS00262: G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;

```


US-09-823-101-3/rev x OAH3_HUMAN ..

Align seg 1/1 to: OAH3_HUMAN from: 1 to: 316

```

537 TTTTGTGTTTCTACGTCCTCTATTATCC...CCCTTAA 500
|||||:|||||:|||||:|||||:|||||:
12 PheileLeuSerGlyPheSerAlaPheProGlnGlnLeuLeuProVal.. 27
499 AAAATAAATTACTCTCTTTTCAATATGTAATCTCCTGCTAAT 450
|||||:|||||:|||||:|||||:|||||:
28 .....LeuPheLeuLeuTyrLeu..... 33
449 AAAAAAAAAAATACACCGGCGATGTTTTTGTATTTCTGTTTCTGTTT 400
|||||:|||||:|||||:|||||:|||||:
34 .....LeuMetPheLeuPheThrLeu 40
399 .....ATCCA 395
41 LeuGlyAsnLeuLeuLeuLeuMetaIaThrValTPrIleGluArgArgLeuH 57
394 TCTTCTGTTGATATTTCTGTCGTCCTCTCTCTCTCTCTCTCTCTCTAT 345
|||||:|||||:|||||:|||||:|||||:
57 sThrProMetTyrLeuPheLeuCysAlaLeuSerIleSerGluIleLeuP 74
344 ATAAATATCTCTGTTTAAATGCGTGGGATGATGATGATGATGATGATG 295
|||||:|||||:|||||:|||||:|||||:
74 heThrValAlaIleThrProArgMetLeuAlaAspLeuLeuPhe..... 88
294 AAACATCATGCCCCCATCGGGGATAGCTCCNCACANGCATCATGTCG 245
|||||:|||||:|||||:|||||:|||||:
89 ...ThrHisArgSerIleThrPheValAla.....CysAl 99
244 TATA.....CTCTGTTCAACACANGTGGACACACAAAGTTGTCCTCA 201
|||||:|||||:|||||:|||||:|||||:
99 alIeGlnMetPhePheSerPheMetPheGlyPheThrHisSerPheLeuH 116
200 TCTTGGTT.....GACACGTCATCGTGAATGATGCCA...CAC 163
|||||:|||||:|||||:|||||:|||||:
116 eMetValMetGlyTyrAspHisTyrValThrIleCysHisProLeuH 132
162 GTT.....CCAGCACTTGTGTCGTCATGATGATGATGATGATGATG 128
|||||:|||||:|||||:|||||:|||||:
133 TyrAsnMetLeuMetSerProArgGlyCysAlaHisLeuValAlaTPrH 149
127 T 127
149 r 149
seq_name: SwissProt_39:ENT1_HUMAN
seq_documentation_block:
ID ENT1_HUMAN STANDARD; PRT; 455 AA.
AC 099808:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (EQUILIBRATIVE
DE NITROBENZYLMECAPTOPURINE RIBOSIDE-SENSITIVE NUCLEOSIDE TRANSPORTER)
DE (EQUILIBRATIVE NBMPR-SENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE
DE TRANSPORTER, ES-TYPE).
GN SLC29A1 OR ENT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.
RP TISSUE=Placenta;
RX MEDLINE=97140266; PubMed=986748;
RA Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E.,
RA Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;
RT "Cloning of a human nucleoside transporter implicated in the cellular
uptake of adenosine and chemotherapeutic drugs."

```

```

RL Nat. Med. 3:89-93(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine, and Jejunum;
RA Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;
RT "Critical structural determinants for high affinity binding of
RT nucleosides to the equilibrative NBMPR-sensitive nucleoside
RT transporter (es) cloned from the human jejunum."
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO
CC LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE
CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER
CC AFFINITY FOR ADENOSINE, INHIBITED BY DIPIRIDAMOLE AND DILAZEP
CC (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND,
CC ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U81375; AAC51103.1; -.
DR EMBL: AF079117; AAC62495.1; -.
DR MIM: 602193; -.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran: 2.
DR PRINTS: PR01130; DERENTRNSRPT.
DR ProDom: PD005103; DER_Nucleoside_tran: 1.
KW Transmembrane; Transport; Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 12 28 POTENTIAL.
FT DOMAIN 29 81 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 82 106 POTENTIAL.
FT DOMAIN 107 110 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 111 129 POTENTIAL.
FT DOMAIN 130 137 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 138 156 POTENTIAL.
FT DOMAIN 157 173 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 174 198 POTENTIAL.
FT DOMAIN 199 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 226 POTENTIAL.
FT DOMAIN 227 290 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 291 310 POTENTIAL.
FT DOMAIN 311 322 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 323 341 POTENTIAL.
FT DOMAIN 342 358 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 359 377 POTENTIAL.
FT DOMAIN 378 392 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 393 412 POTENTIAL.
FT DOMAIN 413 430 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 431 451 POTENTIAL.
FT DOMAIN 452 455 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 47 N-LINKED (GLCNAC...)(POTENTIAL).
SQ SEQUENCE 455 AA; 50088 MW; 9098B95E26515850 CRC64;

```

alignment_scores: Quality: 73.50 Length: 235
Ratio: 0.758 Gaps: 14
Percent Similarity: 41.277 Percent Identity: 22.553

alignment_block: US-09-823-101-3/rev x ENT1_HUMAN ..

Align seg 1/1 to: ENTL_HUMAN from: 1 to: 455

```

648 ACTAAATATTTCTCTCGCGCTCCCTCTCATCTGCAATATACTATA 599
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
35  ThrlIntyPhetHrAsnArgLeuAspMet.....SerGlnAs 47
598 TTATCTCTCTGTC.....GTGT 582
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
47  nValSerLeuValThrAlaGluLeuSerLysAspAlaGlnAlaSerAla 64
581 CTCCTCCCTCCGCC...CCGACCAAAAAAATGTTTGTCTCTTAATTT 535
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
64  lArProAlaAlaProLeuProGluArgAsnSerLeuSerAlaIlePheAsn 80
534 TTGTTTCTACTCTCTCTCTTATTCGCCCTTTAAATTAATTAATCT 485
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
81  AsnValMetThrLeuGlySalMetLeuProLeu.....Le 92
484 CTCTCTTTCATATTGTAAATCTCTCTCTGCTAATAAAAAAATAACC 435
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
92  uLeuPheThrTyrLeu.....AsnSerP 100
434 ACGGCGAT..... 427
      |||:|||||:|||||:|||||:|||||:|||||:
100  heLeuHISGlnArgIleProGlnSerValArgIleLeuGlySerLeuVal 116
426 ...GTCTTTTGTATTTCTGTTTCTGTTATTCATCTCTCTGTGTGATA 380
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
117  AlalLeuLeuValPheLeuIleThrAlaIle..... 127
379 TTTCCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 333
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ...LeuValLysValGlnLeuAspAlaLeuProPheValIleThrM 143
332 TTGTT.....ATTAATGTGGGGT.....GGATAA 307
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
143  etIleLysIleValLeuIleAsnSerPheGlyAlaIleLeuGlnGlySer 159
306 ATTTTGGGGGAAAACTCATGTCCCCCATTCG..... 274
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
160  LeuPheGlyLeuAlaGlyLeuLeuProAlaSerLysThrAlaProIleMe 176
273 .....GGGATAGCTCCNCACANGTCACGTTGCTATACTGTT 234
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
176  tSerGlyGlnGlyLeuAlaGlyPhePheAlaSerValAlaMetIleCysA 193
233 CAACCCANTGTGGACAC..... 217
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
193  lAlleAlaSerGlySerGluLeuSerGluSerAlaPheGlyTyrPheIle 209
216 ACAAGTTGCTCTCATGTTGGTTGACACGTCATCGTGAACATGATGCC 167
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
210  ThrAlaCysAlaValIleIle.....LeuThrIleIleCysTyrLe 223
166 ACACGTTCCACGACCTTGTTGCTGTGCATGATATACGTTGAGAATT 117
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
223  uGlyLeuProArg.....LeuGluPhe 231
116 ATAGG 112
231 YIAR 232
seq_name: SwissProt_39:OATP_RAT
seq_documentation_block:
ID OATP_RAT STANDARD: PRT: 670 AA.
AC P46720.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SODIUM-INDEPENDENT ORGANIC ANION TRANSPORTER (ORGANIC ANION
DE TRANSPORTING POLYPEPTIDE).
GN SLC21A3 OR OATP.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=94105118; PubMed=8278353;
RA Jacquemin E., Hagenbuch B., Stieger B., Wolkoef A.W., Meier P.J.;
RT Expression cloning of a rat liver Na(+)-independent organic anion
RT transporter.
RT Proc. Natl. Acad. Sci. U.S.A. 91:133-137(1994).
CC -1- FUNCTION: MEDIATES THE NA(+)-INDEPENDENT TRANSPORT OF ORGANIC
CC ANIONS SUCH AS BROMOSULFOBROMOPHTHALEIN (BSF) AND CONJUGATED
CC (TAUROCHOLATE) AND UNCONJUGATED (CHOLATE) BILE ACIDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: A HIGH LEVEL EXPRESSION IS SEEN IN THE LIVER
CC AND KIDNEY, WHILE A LOWER LEVEL EXPRESSION IS SEEN IN THE BRAIN,
CC LUNG, SKELETAL MUSCLE AND PROXIMAL COLON.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL; LI9031; AAA16451.1; -
CC DR InterPro: IPR002350; Kazal.
CC DR SMART: SM00280; KAZAL.1.
CC KW Transmembrane; Transport; Glycoprotein.
CC FT TRANSMEM 21..38 POTENTIAL.
CC FT TRANSMEM 86..106 POTENTIAL.
CC FT TRANSMEM 156..176 POTENTIAL.
CC FT TRANSMEM 194..214 POTENTIAL.
CC FT TRANSMEM 246..266 POTENTIAL.
CC FT TRANSMEM 313..333 POTENTIAL.
CC FT TRANSMEM 356..376 POTENTIAL.
CC FT TRANSMEM 388..408 POTENTIAL.
CC FT TRANSMEM 515..535 POTENTIAL.
CC FT TRANSMEM 571..571 POTENTIAL.
CC FT TRANSMEM 602..622 POTENTIAL.
CC FT CARBOHYD 62..62 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 124..124 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 135..135 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 492..492 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 670 AA; 74178 MW; 3D53A4E8E1E21536 CRC64;

```

```

alignment_scores:
Quality: 73.50 Length: 144
Ratio: 0.967 Gaps: 9
Percent Similarity: 52.778 Percent Identity: 26.389

```

alignment_block:

US-09-823-101-3/rev x OATP_RAT ..

Align seg 1/1 to: OATP_RAT from: 1 to: 670

```

595 TTCTCTGTCGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 548
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
332 PheIleAsnLysPheThrPheLeuProLysTyrLeuGlnGlnGlnTyrG 348
547 .....TGTTCTTTAAATTTTGTGTTTC.....TACTCTCTCTTA 514
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
348 YLysSerThrAlaGlnAlaIlePheLeuIleGlyValTyrSerLeu... 363
513 TTCCTCCCTCTTTAAATTAATTAATTAATTAATTAATTAATTAATC 464
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
364 .....ProPheIleCysLeuGlnGlyTyr...LeuIleGlyGlyPheIleMe 377

```

```

463 TCCTCTCGCTATATAAAAAAAAAATACACCGCGCATGTTTGTGTAT 414
      :::::  |||  |||  :::::|||||:::
378 LysLysrPhelYsILeThrValLysLysAlaLaIATyLLeuAlaIaPheYsLe 394
      |||||:::  |||  |||||:::|||||:::
413 TTCTGTTTTGTTATGCATCTTCCTGTTGGATATTTCTCTGCCGTC 364
      |||||:::  |||  |||||:::|||||:::
394 uSerValaIaPheGLuTY.....LeuLeu.PheLeuYsHisPhe 406
      :::::|||||:::
363 TCTCTCCCCCCCCCTCATATATAAATATCTCTGTATTATATGCGGGG 314
      |||  :::  :::::||||:  :::  :::
407 MetLeuThrCysAspAsnAlaIaValaIaGLyLeuThrThSeryTrY 423
313 TCGATTAATTTTGGGGGAAAAACATATCTCCCCCATCGGGGATGTC 264
      |||  :::  |||||:::  :::  |||||:::
423 sGLY.....ValGlnHisGlnLeuHisValGIseryValLeuAla 438
263 CMCACAGTCACGCTGCTGTATACCTGTGTTCAAC..... 229
      ::|||:::  |||||:::
438 sPCysAsnThrAlaGlySer.....CysSerThrAsnThrTrpAsPro 452
228 CANTGTGGACACACAGTTGCTCTGCAG 199
      |||||:::  :::::  |||  |||
453 ValCysGlyAspAsnGLyAlaIaTrpMet 462
seq_name: SwissProt_39:VIPs_RAT

seq_documentation_block:
ID VIPs_RAT STANDARD: PRT: 437 AA.
AC P3500;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VASOACTIVE INTESTINAL POLYPEPTIDE RECEPTOR 2 PRECURSOR (VIP-R-2)
DE (PTUTITARY ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE III RECEPTOR)
DE (PACAP TYPE III RECEPTOR) (PACAP-R-3).
GN VIPR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RX (1)
RX SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY; TISSUE=Olfactory bulb;
RX MEDLINE=94039806; PubMed=8224221;
RX Lutz E.-M., Sheward W.J., West K.M., Morrow J.A., Fink G.,
RX Hammar A.J.;
RA "The VIP2 receptor: molecular characterisation of a cDNA encoding a
RT novel receptor for vasoactive intestinal peptide.";
RT FEBS Lett. 334:3-8(1993).
RL (2)
RL SEQUENCE FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=95080140; PubMed=7988457;
RX Usdin T.B., Bonner T.I., Mezey E.;
RA "Two receptors for vasoactive intestinal polypeptide with similar
RL specificity and complementary distributions.";
RL Endocrinology 135:2662-2680(1994)
RL -1- FUNCTION: THIS IS A RECEPTOR FOR VIP AS WELL AS PACAP-38 AND -27,
      THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
      ACTIVATE ADENYLYL CYCLASE. CAN BE COUPLED TO PHOSPHOLIPASE C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z25885; CAA81104.1; -.

```

DR EMBL: U09631: AAB60459.1; -
DR PIR: S39069: S39069.
DR GCRdb: GCR_1034; -
DR InterPro: IPR000832; GPCR_secretin.
DR InterPro: IPR001879; horm_torptor.
DR Pfam: PF00002; 7tm.2; 1.
DR PRINTS: PR00249; GPCRSECRETIN.
DR PRINTS: PR00491; VASOACTVEPR.
DR PRINTS: PR01155; VIP2RECEPTOR.
DR SMART: SM00008; Hormr; 1.
DR PROSITE: PS00649; G-PROTEIN_RECEP_F2.1; 1.
DR PROSITE: PS00650; G-PROTEIN_RECEP_F2.2; 1.
DR PROSITE: PS50227; G-PROTEIN_RECEP_F2.3; 1.
DR G-protein coupled receptor; transmembrane; glycoprotein; Signal
KW SIGNAL
FT CHAIN 1 22
FT 23 437
FT 23 125
FT TRANSSEM 126 150
FT DOMAIN 151 157
FT TRANSSEM 158 177
FT DOMAIN 178 202
FT TRANSSEM 203 226
FT DOMAIN 227 239
FT TRANSSEM 240 261
FT DOMAIN 262 278
FT TRANSSEM 279 302
FT DOMAIN 303 327
FT TRANSSEM 328 347
FT DOMAIN 348 359
FT TRANSSEM 360 379
FT DOMAIN 380 437
FT CARBOHYD 57 57
FT CARBOHYD 87 87
FT CARBOHYD 91 91
FT CONFLICT 383 383
SO SEQUENCE 437 AA; 49552 MM; 7E10218A9EE31360 CRC64;

alignment_scores:
Quality: 73.00 Length: 131
Ratio: 1.197 Gaps: 6
Percent Similarity: 46.565 Percent Identity: 23.664

alignment_block:
US-09-823-101-3 x VIPS_RAT ..

Align seg 1/1 to: VIPS_RAT from: 1 to: 437

```

57 TGCCTTGGCGTGATCTTCTGTGACTCTCTGCTATGATCCTGTGCTCTAG 106
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
9 CysTYrCysTPrLeuLeuValAlaValSerIleHisProGluCysAr 25
107 CGTATCCATATATTCTCAACAGCTATATCGATGACAGCAGCAAGTCGG 156
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
25 gPheHisLeuGluIleGlnGlu.....GlnGluThrIlyscysA 38
157 TGGACGGTGGGATGAGTCAGATTCACAGTGCAGTCGGTCAACCAATGAGG 206
||||:||||:||||:||||:||||:||||:||||:||||:||||:
38 IagIleuLeuSerIleMetGlu.....Asn 47
207 CACAACCTGTGTGTCACACANTGTTTGAACAGAGTATAGCAACAAGTGAC 256
||||:||||:||||:||||:||||:||||:||||:||||:||||:
48 HisArgAlaCys.....SerGlyValTrpAspAsnIleTh 59
257 NTGTGNGAGCTATCCCGATGGGGGACATGAGTTTTCGCCCAAAAT 306
||||:||||:||||:||||:||||:||||:||||:||||:||||:
59 rCysTrpArpProAlaAspIleGlyGluThValThrValArpGlySPrOL 76
307 TTATCCACCCACCAATTAATTAACAAGAGATATTTATATGAGGGGGGG 356
:::||||:||||:||||:||||:||||:||||:||||:||||:
76 ySvalPheSerAsn.....PheTYrSerArY..... 84

```

357 GGAGAGACGACGACACGAGAAATATCAACAG.....GAGATGG 397
 85ProGlyAsnIleSerIysnScyThrSerArgPgl 96
 398 ATAAACGAAAAACAGAAATACAAAAACACATCGCGGTGGTTAT 440
 96 YTPserGluThrPheProAspPheIleAspAlaCysGlyTyr 110

seq_name: SwissProt_39:08D2_HUMAN

seq_documentation_block: ID 08D2_HUMAN STANDARD; PRT; 311 AA.

AC Q9GZM6;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE OLFACATORY RECEPTOR 8D2 (OLFACATORY RECEPTOR-LIKE PROTEIN JCG2).
 GN OR8D2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tongue;
 RA Gaubin J.-C., Chobert J.-M., Haerle T.;
 RT "Molecular cloning of olfactory receptor-like mRNAs expressed in human tongue."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE TONGUE.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF162668; AAG43386.1; -;
 DR EMBL: AF162669; AAG43387.1; -;
 DR InterPro: IPR000276: GPCR_Rhodopsn.
 DR Pfam: PF00001: 7tm_1.1
 DR PRINTS: PR00237; GPCR_RHODOPSN.
 DR PROSITE: PS00237; G_PROTEIN_RECPT_FL_1; FALSE_NEG.
 DR PROSITE: PS50262; G_PROTEIN_RECPT_FL_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 26 49 1 (POTENTIAL).
 FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 58 79 2 (POTENTIAL).
 FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 120 3 (POTENTIAL).
 FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 140 158 4 (POTENTIAL).
 FT DOMAIN 159 195 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 196 219 5 (POTENTIAL).
 FT DOMAIN 220 236 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 237 259 6 (POTENTIAL).
 FT DOMAIN 260 272 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 273 292 7 (POTENTIAL).
 FT DOMAIN 293 311 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 97 189 BY SIMILARITY.
 FT CARBOHYD 5 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 311 AA: 34857 MW: ECE606FBBC962A82 CRC64;

alignment_scores:

Quality: 72.50 Length: 178
 Ratio: 0.863 Gaps: 11
 Percent Similarity: 47.191 Percent Identity: 24.719

alignment_block:

us-09-823-101-3/rev x 08D2_HUMAN

Align seg 1/1 to: 08D2_HUMAN from: 1 to: 311

696 GGCCTATACCTG.....AGATTATTTTGT 671
 33 GlyIleTyrValAlaThrValAlaGlyAsnIleuGlyMetIlePheLeu 49
 670 TCTTCATTATAGTCGG.....CCCCACATAATTTCTCGGCGTC 627
 49 eAlaIleuSerSerGlnIleuTyrProValTyrTyrPheIleuSerHisL 66
 626 TCCCTTCATAGTCGATTAATACATTAATATTCCTGCTGCTGCC 577
 66 euserPheIleasp.....LeucyTyrSerSerValIleThrPro 79
 576TCCCTCCCCCACCACAAAAATAGTT..... 550
 80 LysMetLeuValAsnPheValProGluGluAsnIleIleSerPheLeuG 96
 549TTTGTTCTTAATTTTGTGTTCTACT.... 523
 96 uCysIleThrGlnIleuTyrPhePheLeuIlePheValIleAlaGluTyr 113
 522CTCTCTTATTCGCCCC 505
 113 yIleuLeuThrAlaMetGluTyrAspArgTyrValAlaIleCysArgPro 129
 504 CTTTAAAAATTAATTAATCTCTTTCATATTGTAATCTGCTGCTGC 455
 130 Leu.LeuTyrAsnIleValMetSerHisArgValCysSerIleMeteta 146
 454 CTAAATAAAAAAATA.....ACCACGCG 429
 146 IValValIleTyrSerLeuGlyPheLeuThrAlaThrValHisThrThrArg 162
 428 ATGTGTTTTTGTATTTCTGT.....TTGCTTATCATCTCTCTGT 385
 163 MetSerValIleuSerPheCysArgSerHisThrValSerHisThrPheC 179
 384 TGAATTTCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 341
 179 sAspIleLeuProLeuLeuThrLeuSerCysSerSerThrHisIleAsnG 196
 340 AATATCTCTGTTATTAATGTTGGGTGAT 309
 196 IuIleLeuPheIleIleGlyGlyValAsn 206

seq_name: SwissProt_39:VE5_RHPV1

seq_documentation_block: ID VE5_RHPV1 STANDARD; PRT; 157 AA.

AC P24834;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROBABLE E5 PROTEIN.
 GN E5.
 OS Rhesus papillomavirus type 1 (Rhpv 1).
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_Taxid=10570;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91135018; PubMed-1847267;
 RA Ostrow R.S., Labresh K.V., Fairs A.J.;
 RT "Characterization of the complete Rhpv 1 genomic sequence and an integration locus from a metastatic tumor."

OC Oestridea; Oestridea; Hypoderma.
 OX NCBI_TaxID=7389;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95107346; PubMed=7808473;
 RA Moire N., Bigot Y., Perleuet G., Boulard C.;
 RT "Sequencing and gene expression of hypodermis A, B, C in larval
 stages of Hypoderma lineatum."
 RL Mol. Biochem. Parasitol. 66:233-240(1994).
 RN [2]
 RP SEQUENCE OF 3-256 FROM N.A.
 RA Kuhn I., Files J.G., Pruett J.H., Temeyer K.B.;
 RL Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 31-51.
 RX MEDLINE=81232922; PubMed=7018579;
 RA Tong N.T., Imhoff J.M., Lecroisey A., Kell B.;
 RT "Hypodermin A, a trypsin-like neutral proteinase from the insect
 Hypoderma lineatum."
 RL Biochim. Biophys. Acta 658:209-219(1981).
 CC -1- FUNCTION: SPECIFICITY, LIMITED TO CARBOXYL SIDE OF ARGININE
 CC RESIDUE IN B-CHAIN OF INSULIN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DEVELOPMENTAL STAGE: LARVAL-SPECIFIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X74303; CA552356.1; -;
 DR EMBL: L24814; AA829224.1; -;
 DR PIR: A21590; A21590.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.111; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR00089; Trypsin.
 DR Pfam: PF00089; Trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00200; TRYP_SPE.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS.1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 KW Hydrolyase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 30
 FT CHAIN 31 256
 FT ACT_SITE 71 71
 FT ACT_SITE 116 116
 FT ACT_SITE 210 210
 FT DISULFID 56 72
 FT DISULFID 180 197
 FT DISULFID 206 230
 FT SITE 204 204
 FT SITE 37 37
 FT CONFLICT 45 45
 FT CONFLICT 51 51
 FT CONFLICT 248 248
 SO SEQUENCE 256 AA; 27907 MW; 63D12D337F66AA61 CRC64;
 alignment_scores:
 Quality: 71.50 Length: 87
 Ratio: 1.833 Gaps: 4
 Percent Similarity: 44.828 Percent Identity: 24.138
 alignment_block:
 US-09-823-101-3/rev x HYPA_HYPLI ..

Align seg 1/1 to: HYPA_HYPLI from: 1 to: 256
 354 CCCCCCATATATAATATCTCTTTATTATTGTTGGGTGATTAAT 305
 |||||:::|||||:::|||||:::|||||
 142 PROPRGIIITyTAlaAspAlaIleValSerGlyTP..... 153
 304 TTTTGGGGAANAACATCATGCCCCATCGGAGATAGTCGN..... 262
 |||||:::|||||:::|||||:::|||||
 154GlyGIuThrLeuValProProGlyIleProAspGlnLeu 168
 261CACANGTCAGTGTCCT..... 244
 168 rGSeValAspVallylleIleHisArgIuIuScysAlaSerArgasn 184
 243ATACCTGTTCAACACANG 224
 185 pHeGlyTyTylSerAsnIleLysAlaSerMetIleCysAlaTyrAlaI 201
 223 TGGACACACACACTTGT.....GCTCATGTTGGTTGACC 189
 |||||:::|||||:::|||||:::|||||
 201 eGlyLysAspSerCysGlnGlyAspSerCylGlyProLeuValAsna 218
 188 ACCTCATCTGTG 178
 :::::|||||
 218 snLeuLeuVal 221
 seq_name: SwissProt_39:YKR9_YEAST
 seq_documentation_block:
 ID YKR9_YEAST STANDARD; PRT; 679 AA.
 AC P34237; P34236;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-JUN-1994 (Rel. 1, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHEtical 77.5 KDA PROTEIN IN RPL17A-STE3 INTERGENIC REGION.
 GN YKL179C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205264; PubMed=8154185;
 RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
 RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
 RA Banerji A., Ansoerge W.;
 RT "Sequencing and analysis of 51.6 kilobases on the left arm of
 RT chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
 RT frames including the FAS1 gene."
 RL Yeast 9:1343-1348(1993).
 RN [2]
 RP REVISIONS.
 RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
 RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
 RA Banerji A., Ansoerge W.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME TO HUMAN CENP-E, AND TO YEAST USO1 AND S. POMBE
 CC SPC364.04C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X74151; CA552259.1; -;
 DR EMBL: Z28179; CA82021.1; -;
 DR PIR: S38011; S38011.
 DR SGD: S0001662; YKL179C.
 KW Hypothetical protein.

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-3 to: SPTREMBL_17:* out_format : pfs
Date: Jan 17, 2002 4:07 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framer.n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09823101/runat_17012002.155138.22043/app-query.fasta.1.2239
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=n2p.rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OCAPOP=4.500 -OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=Pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USPR=US09823101.ecgn1_1_384 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEX
-WAIT -THRADE=1

Search information block:

Query: US-09-823-101-3
Query length: 698
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 224.230000

```
score_list:
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
SP_Invertebrate:018560 - 90.50 160.08 0.2185 555 018560 vargula hilgendorffii (
SP_Invertebrate:026195 - 88.50 163.25 0.4041 200 026195 plasmodium vivax, pval
SP_plant:Q9M1P9 - 86.00 153.93 0.7162 373 09M1P9 arabidopsis thaliana (mc
SP_Invertebrate:09XWS9 - 85.00 151.52 0.9183 366 09XWS9 caenorhabditis elegans
SP_Invertebrate:Q9N8H0 - 82.50 136.22 1.49 1737 09N8H0 trypanosoma brucei, f
SP_plant:Q9AW73 - 82.00 144.99 1.96 430 09AW73 guillardia theta (cryptic
SP_plant:Q9LR94 - 82.00 142.07 1.87 753 09LR94 arabidopsis thaliana (mc
SP_bacteria:051099 - 82.00 140.89 1.84 773 051099 borrelia burgdorferi (ly
SP_plant:Q9ZS19 - 81.00 140.09 2.42 652 09ZS19 arabidopsis thaliana (mc
SP_virus:Q91BR4 - 80.00 140.51 3.23 462 091BR4 spodioplera exigua nucleoc
SP_bacteria:0950U3 - 77.50 136.11 6.16 426 0950U3 shigella sonnei. orfas.
SP_Invertebrate:Q9F740 - 77.50 136.11 6.16 426 09F740 shigella sonnei. wzx. 6/
SP_Invertebrate:Q9NG78 - 77.50 131.79 5.78 791 09NG78 caenorhabditis elegans
SP_Invertebrate:062186 - 77.00 135.85 7.07 384 062186 caenorhabditis elegans
SP_fungi:Q9P472 - 77.00 134.40 6.92 472 09P472 emeticella nidulans (ast
SP_plant:Q9PFU7 - 76.50 134.45 7.98 407 09PFU7 arabidopsis thaliana (mc
SP_bacteria:035044 - 76.50 134.13 7.94 426 035044 shigella sonnei. form 1
SP_virus:Q9DMW2 - 76.00 137.72 9.66 221 09DMW2 rat cytomegalovirus (str
SP_plant:Q9FRV9 - 76.00 136.92 9.54 348 09FRV9 ruminococcus flavifaciens
SP_bacteria:09S521 - 76.00 134.51 9.21 350 09S521 escherichia coli. putati
SP_rodent:Q9R1V7 - 76.00 128.48 8.42 829 09R1V7 mus musculus (mouse). ad
SP_plant:Q9SC36 - 75.50 137.22 11.05 206 09SC36 pisum sativum (garden pe
SP_Invertebrate:Q9VD26 - 75.50 134.62 10.63 299 09VD26 drosophila melanogaster
SP_bacteria:Q9H0Y3 - 75.00 131.22 10.26 421 09H0Y3 plesiomonas aeruginosa.
SP_bacteria:Q9HMT2 - 75.00 131.27 11.66 471 09HMT2 pseudomonas aeruginosa.
SP_fungi:Q9HGU6 - 75.00 130.45 11.52 471 09HGU6 blumeria graminis. prote
SP_fungi:Q9H077 - 75.00 130.32 11.50 480 09H077 aspergillus niger. camp
SP_bacteria:084150 - 75.00 129.93 11.43 507 084150 chlamydia trachomatis. m
SP_virus:Q9WS53 - 75.00 123.61 10.73 933 09WS53 oryza sativa (rice). hyfr
SP_human:Q14674 - 75.00 121.09 10.44 1217 09WS53 simian t-cell lymphotr
SP_fungi:Q9UUS9 - 74.50 129.44 13.03 472 09UUS9 erysipe graminis (subst
SP_bacteria:Q9PKP0 - 74.50 128.95 12.99 506 09PKP0 chlamydia muridarum. mon
SP_plant:Q9STN1 - 74.00 130.34 15.33 350 09STN1 arabidopsis thaliana (mc
SP_plant:Q94532 - 74.00 125.12 14.14 760 094532 arabidopsis thaliana (mc
SP_bacteria:Q913K3 - 74.00 124.83 14.08 792 0913K3 ruminococcus flavifaciens
SP_rodent:Q9EPK5 - 73.50 130.30 17.60 314 09EPK5 mus musculus (mouse). b2
SP_Invertebrate:Q9N477 - 73.50 129.96 17.51 330 09N477 caenorhabditis elegans
SP_Invertebrate:Q9GR42 - 73.50 129.37 17.36 359 09GR42 toxoplasma gondii. mid
SP_Invertebrate:Q44350 - 73.50 128.82 17.22 388 044350 manduca sexta (tobacco
```

```
SP_bacteria:Q91347 - 73.50 128.52 17.15 405 031347 bacillus cereus. orfi
SP_human:Q9UJY2 - 73.50 127.69 16.94 456 09UJY2 homo sapiens (human).
SP_human:Q9H6L0 - 73.50 126.50 16.64 541 09H6L0 homo sapiens (human).
SP_plant:Q9SKD1 - 73.50 125.10 16.30 661 09SKD1 arabidopsis thaliana
SP_human:Q9Y4G1 + 73.50 122.23 15.62 997 09Y4G1 homo sapiens (human).
```

seq_name: sp_invertebrate:018560

seq_documentation_block:

```
ID 018560 PRELIMINARY; PRT; 555 AA.
AC 018560:
DT 01-JAN-1998 (TRMBREL. 05 Created)
DT 01-JAN-1998 (TRMBREL. 05, last sequence update)
DT 01-JUN-2001 (TRMBREL. 17, last annotation update)
DE LUCIFERASE.
OC Vargula hilgendorffii (Sea firefly) (Cypridina hilgendorffii).
OC Eukaryota: Metazoa: Arthropoda: Crustacea: Ostracoda: Myodocopa:
OC Myodocopidae: Cypridinoidae, Cypridinidae, Vargula.
CX NCBI_TaxID=6674;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250550; PubMed=9096406;
RA Miesenböck G., Rothman J.E.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89490; AAB86460.1; -.
DR InterPro; IPR001846; Vwd.
DR Pfam; PF00094; vwd; 2.
DR SMART; SM00216; VWD; 2.
SQ SEQUENCE 555 AA; 61583 MW; CE12E56060877863 CRC64;
```

alignment_scores:

```
Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6
Percent Similarity: 38.060 Percent Identity: 28.358
```

alignment_block:
US-09-823-101-3/rev x 018560 ..

Align seg 1/1 to: 018560 from: 1 to: 555

```
372 TGTCCGTCCTCTCCCTCCCTCCCTCATATAATATCTTGTATTAA 323
|||||
22 CysProvaIgluAgluProProSer..... 30
322 TTGTGTCGGGTGATAAATTTTGGGAAACATCATGCCCCCATCG 273
|||||
31 .....SerThrProThrValProThrSerC 39
272 GGATAGCTCCNCAC.....ANGTCAGTTGTGCTATA..... 241
|||
39 ysgluAlaIySgluIyLucIySleAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTTCAACACANTGTGGACACACAG 212
56 ArgAspIleLeuSerAspIyLeuCysGluAsnLysProGlyLysThrCys 72
211 TTGTGTCATCATGTTGTTGACACAGTCATCTGTAACATGCACACAG 162
|||
72 scYs..... 73
161 TTCACGACATCTGTCATCATGATGATGATGATGATGATGATGATG 112
|||
74 .....ArgMetCysGlnTyValIleGluCysArgValGluAlaAlaG 88
111 ATACGCTAGACACAGATGATGATGATGATGATGATGATGATGATG 62
|||
```


452ATAAAAAAAAAATACACGGCATGCTTTTGTAT 414
 225 leasnlystleuasnarglulysilleuPhepkelelleuLeuLyr 241
 413 TCTCTGTTTCGTTATCCATCTTCCTGTGATATTCTCGTCCGTC 364
 242lleuasnleuPhe..... 245
 363 TCTCTCCCCCCCCCATATATAATATCTGTTATTATTGTTGGG 314
 246SerProPheSerThrArgllelleuLeuLysPhe.....L 259
 313 TGGATAATTTTTGGGGAATACTATGTCCTCCCATCGGGATAGTC 264
 259 euasnlystleuLysasnAsnleuLys.....llealallelle 273
 263 CACACAGTCACGTTGCTGCTATCTGTTCAACCACTGTCACACA 214
 274 LeullearghIstYrIlePheYrSerLysGlnThrAsnIleasnIas 290
 213 AGTGTG 207
 290 292

seq_name: sp_plant:09LR94

seq_documentation_block:

ID 09LR94 PRELIMINARY; PRT: 653 AA.

AC 09LR94;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE T23E23.16
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C.,
 RA Khan S., Kim C., Alcafi H., Bei B., Chin C., Chio J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thuerl A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Thuelogis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C.,
 RA Khan S., Kim C., Alcafi H., Bei B., Chin C., Chio J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thuerl A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Thuelogis A., Ecker J.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 TRANSCRIPTASE)
 CC EMBL: AC002423; AF87143.1; -
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; Irc; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 653 AA; 75802 MW; D7AEA9219C604814 CRC64;

alignment_scores:

Quality: 82.00 Length: 180
 Ratio: 0.965 Gaps: 11
 Percent Similarity: 47.222 Percent Identity: 23.889

alignment_block:

US-09-823-101-3 x 09LR94

Align seg 1/1 to: 09LR94 from: 1 to: 653

11 CGTCCACGGG.....TTCCGCTGTGACCCCTATGCTGTCACAG 54
 511 leuSerThrGlyAspLysMetLeuGlnTrpAsnArgAlaLeuSerProth 527
 55 T...GTGCTTTCCTGGAT.....CTTCTGTGA 80
 527 rcysValleuCySAsnAsnAsnIleGlnThrArgAsnHisleuPhePhe 544
 81 GTCCTGTATGATCCCTGCTCCTAGCGTATCTATTAATTCACAGTC 130
 544 ecyscystYrThrAla..... 549
 131 ATATGATGACAGACAGACAGAGCGGTGAGACGTGTCACAGTTCAC 180
 550 ...GlnIleTrpGlnLeuAlaLysAsnIleTYrLysAlaLysPhe 565
 181 GATGACGTGTCACACACATGAGACACACTGTGTCACACANTG. 229
 565 rhrAsnTrpSerThrIleLeuThrSer.....ValSerThrThrTPA 580
 230TTGACAGAGTATAGCACAAAGTGACNTGTGNGAGACTATC 271
 580 rgAsnArgThrGlnSerPheLeuAlaArgTYr..... 590
 272 CCCGATGGGGGACATGATTTTCCCAAAATTTATCCACCCACCA 321
 591llePheGlnAlaThrIleHis..... 597
 322 ATTAATAACAGATATTTTATGAGCGGGGGAGAGACAGCGAC 371
 598ThrIleTrpHisGlnIleArgAsnGlyArgArgHisGlyG 610
 372 ACAGAAATATCAACAGAGAGA.....TGATAAACAAGAAC... 409
 610 lArgSerAsnSerAlaThrHisleuIleTrpTrpLeuAspLysGlnMet 626
 410 AGAAATACAAAACACATCGCGTGTATTTTATTATTAAGCAG 459
 627 ArgAsnGlnIleSerThrIleAla.....Alase 636
 460 AGCAGATTACAATATGAAGAAGAGAGACTAATTTATTT 499
 636 rglAspHisArgTYrAspLysProleuGlnLeuTrpPhe 649
 seq_name: sp_bacteria:051099
 seq_documentation_block:
 ID 051099 PRELIMINARY; PRT: 773 AA.
 AC 051099;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE HYPOTHETICAL 89.8 KDA PROTEIN.
 GN B80072.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,

RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utechtack T., Wathley L., McDonald L., Artach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic Sequence of a Lyme disease spirochaete, Borrelia
 burgdorferi.";
 RL Nature 390:580-586(1997).
 DR EMBL; AE001120; AAC66467.1; -;
 DR TIGR; BB0072; -;
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 773 AA; 89847 MW; 42M49219431EF362 CRC64;

alignment_scores:
 Quality: 82.00 Length: 135
 Ratio: 1.155 Gaps: 7
 Percent Similarity: 52.593 Percent Identity: 28.889

alignment_block:
 US-09-823-101-3/rev x 051099 ..

Align seg 1/1 to: 051099 from: 1 to: 773

```

680 TATTTTGTTCCTTCATTATAGTCGCCCCCACTAAATATTTCTCTCGG 631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 TyrPheAlaIleIleLeuSerLeuSerProSerLeuIlePheSerPhe.. 440
630 CGTCGCCCTCATAGTCGCAATACATATATTTCTCTCGTCGTCGTC 581
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 .....LeuAsnAsnThrLysPheMetThrAspThr.... 450
580 TCCTCTCTCCCCCCCCACCAAAATAGTTTTCCTTTAATTTTGT 531
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 .....LeuLeuLeuIlePheIle 456
530 TTTCTACTCTCTCTTATTCCTCCCTTAAATAATATCTCTCTC 481
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
456 eSerValAlaIleIleIleProIleMetLeuLysIleAsnThrLysCysL 473
480 TTTTCATATTTGTAT.....CTCTCTCTGCTTAT 450
|||:|||||:|||||:|||||:|||||:|||||:|||||:
473 euserPhePheValTyrAlaIleLeuMetIleSerIleIleLeuProLeu 489
449 AAAAAAATAATACC.....ACGGCATGTG 424
:::|||||:|||||:|||||:|||||:|||||:|||||:
490 AspleuGluIleThrLeuLysIleIleSerLeuLeuPheThrPheLeuVa 506
423 TTTTGTATTTCTGTTTGTTCATTCATCTCTCTGTGATATTTCT 374
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 lPhePheTyrPhe..PheArgLeuSerAlaPheLeuAlaSn..... 519
373 GTGTCGGCTCTCTCCCCCCCCCTCATATAAATATCTCTGTATTA 324
|||:|||||:|||||:|||||:|||||:|||||:|||||:
520 .....ProIleLeuIleSerPheMetPheLeuVal 529
323 ATT 321
530 Leu 530
```

seq_name: sp_plant:Q9ZS19

seq_documentation_block:
 ID Q9ZS19; PRELIMINARY; PRT; 652 AA.

AC Q9ZS19;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
 DE T15B16.10 PROTEIN (PUTATIVE CHP-RICH ZINC FINGER PROTEIN).
 GN T15B16.10 OR AT4G01740.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA WASHU;
 RT "The A. thaliana Genome Sequencing Project.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Stoneking T., Smith R.;
 RT "The sequence of A. thaliana T15B16.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Waterston R.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF104919; AAC72874.1; -;
 DR EMBL; AL161492; CAB77744.1; -;
 DR InterPro; IPR001841; Znf_finger.
 DR PROSITE; PS00518; ZINC_FINGER_C3HC4; UNKNOWN.1.
 SQ SEQUENCE 652 AA; 74609 MW; DB4190528145D259 CRC64;

alignment_scores:
 Quality: 81.00 Length: 209
 Ratio: 0.931 Gaps: 17
 Percent Similarity: 41.627 Percent Identity: 25.837

alignment_block:
 US-09-823-101-3/rev x Q9ZS19 ..

Align seg 1/1 to: Q9ZS19 from: 1 to: 652

```

584 TGTCCTCTCTCCCCCCCCACCA..... 561
|||:|||||:|||||:|||||:|||||:|||||:|||||:
105 CysAlaLysTyrProProGluValIleAsnIleSerLysThrHisH 121
560 .....AAAAATAGTTTGTCTTAAATATTTGTTTC 527
|||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ShiLysLeuAsnLeuLeuLysGluLeuIleGlnPheAspTyrCysVal. 137
526 TACTCTCTCTTATTCCTCCCTTAAATAATATCTCTC..... 483
:::|||||:|||||:|||||:|||||:|||||:|||||:
138 .....TrpHisProProGluValAsnHisThrLeuGluValAsn 150
482 ...TCTTTCAATATTTGTATCTCTC.....CTGCC 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 HisSerTyrHisSerLeuHisProLeuLysLeuHisThrGlyGlnLeuPr 167
453 TAAATAAATAAATAATACACGCCCATCTCT..... 423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 oAsnTyrSerAspArgLysCysArgLeuCysAlaLysGluIleGluValG 184
422 ..TTTGTATTTCTGTTTGTTCATTCATCTCTCTGTGATATTTCC 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 lLysPheThrHisCys...SerLeuCysAsnPheThrLeuAspMetSer 199
374 TGTCGCCCTCTCTCCCCCCCCCTCATATAAATATCTC..... 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 CysVal.....LeuAsnProGln.....ArgTyrLeuTyrAsnLe 212
332 .....TTGTATTAATTTGTGCGGGGTGGATTAATTTT 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 uLysAlaHisAspHisGlnLeuThrLeuProSerLeuValArgSerHel 229
301 TGGGGAAAACTCATGTCCCGCCATCGGGGATAGTCNCACANGTCAC 252
```



```

93 PheMetPheTyrSerTyrValSerAspSerIleSerIleuThrLeuVal 109
627 CTCCCTCATAGCTGCATATACTATATATCTCTCGCGTCCTCC 578
109 IYrIleSer.SerAlaTyrPheValCysIleSerSerGlyArgPheSer 125
577 CTCTCCCGCCCGCCACCAAAA.....AATAGTTT 549
126 LeuLeuGlnAlaValAlaGlyArgPheArgCysGluLeuTyrIleAsnIleTyr 142
548 TTGTTCTTAATTTTGT.....TTCTACTCTCTCTTATTC 511
142 rSerThrIleIleTyrIleGlyCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAATAATTAATCTCTCTTTTCATATTGTATCTCC 461
159 IuPro.....LeuTyrTyrSer..... 164
460 TCCTGCCTAATATAAAAAAATAAACACGCGCATGTGTTTTGTATTTC 411
165 .....AlaIleSe 167
410 TGTTCCTTATCCATCTCTCTGTGATATTCTGTGTCGCTCTCT 361
167 rIlePheIleTyrSerIleSerIleuValPheSerSerHisLysCysA 184
360 CTCCCGCCCGCCCT.CATATAAA..... 339
184 snValProCysPheHisIleLysArgProSerIleLeuValTyrLysAsp 200
338 .....TATCTCTGTTATTAT 322
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValValle 217
321 TGTGGGGGTGATATAATTTTGGGGGAAAACTCATGTCCCCCATCGGG 272
217 uSerSerIleAspLeuPheIleLeuLysGluTyrPheSerTyrAsn.... 232
271 GATAGCTCCNACANGTCAGCTGCTATCTGCTCAACACANGTG 222
233 .....SerValAlaIleTyrGlnValValThrArgVal 243
221 GACACACAGTGTGCTC 204
244 AsnThrGlyLeuIleIle 249

seq_name: sp.bacteria:Q9F740

seq_documentation_block:
ID Q9F740 PRELIMINARY; PRT; 426 AA.
AC Q9F740;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE W2X.
GN W2X.
OS Shigella sonnei.
OG plasmid pInv.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCB1_TaxID=624;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53G1.
RX PubMed=10992522;
RA Shepherd J., Wang L., Reeves P.R.;
RT "Comparison of the O antigen gene clusters of Escherichia coli
RT (Shigella) Sonnei and Plesiomonas shigelloides O17: Sonnei gained its
RT current plasmid borne O antigen genes from Plesiomonas shigelloides in
RT a recent event.";
RL Infect. Immun. 68:6056-6061(2000).
DR EMBL; AF285971; AAG17420.1; -.

```

```

DR InterPro: IPR002797; Polysacc_synt.
DR Pfam: PF01943; Polysacc_synt; 1.
KW Plasmid.
SQ SEQUENCE 426 AA: 49279 MW: C5E28DCC21038E5 CRC64;

alignment_scores:
Quality: 77.50 Length: 190
Ratio: 0.891 Gaps: 7
Percent Similarity: 45.789 Percent Identity: 22.632

alignment_block:
US-09-823-101-3/rev x Q9F740 ..
Align seg 1/1 to: Q9F740 from: 1 to: 426

677 TTTTGTTCCTCATTAATAGTGGCCCCCACTAATATTTCTCGGCGT 628
93 PheMetPheTyrSerTyrValSerAspSerIleSerIleuThrLeuVal 109
627 CTCCCTCATAGCTGCATATACTATATATCTCTCGCGTCCTCC 578
109 IYrIleSer.SerAlaTyrPheValCysIleSerSerGlyArgPheSer 125
577 CTCTCCCGCCCGCCACCAAAA.....AATAGTTT 549
126 LeuLeuGlnAlaValAlaGlyArgPheArgCysGluLeuTyrIleAsnIleTyr 142
548 TTGTTCTTAATTTTGT.....TTCTACTCTCTCTTATTC 511
142 rSerThrIleIleTyrIleGlyCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAATAATTAATCTCTCTTTTCATATTGTATCTCC 461
159 IuPro.....LeuTyrTyrSer..... 164
460 TCCTGCCTAATATAAAAAAATAAACACGCGCATGTGTTTTGTATTTC 411
165 .....AlaIleSe 167
410 TGTTCCTTATCCATCTCTCTGTGATATTCTGTGTCGCTCTCT 361
167 rIlePheIleTyrSerIleSerIleuValPheSerSerHisLysCysA 184
360 CTCCCGCCCGCCCT.CATATAAA..... 339
184 snValProCysPheHisIleLysArgProSerIleLeuValTyrLysAsp 200
338 .....TATCTCTGTTATTAT 322
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValValle 217
321 TGTGGGGGTGATATAATTTTGGGGGAAAACTCATGTCCCCCATCGGG 272
217 uSerSerIleAspLeuPheIleLeuLysGluTyrPheSerTyrAsn.... 232
271 GATAGCTCCNACANGTCAGCTGCTATCTGCTCAACACANGTG 222
233 .....SerValAlaIleTyrGlnValValThrArgVal 243
221 GACACACAGTGTGCTC 204
244 AsnThrGlyLeuIleIle 249

seq_name: sp.invertebrate:Q9NG78

seq_documentation_block:
ID Q9NG78 PRELIMINARY; PRT; 791 AA.
AC Q9NG78;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HERG-LIKE POTASSIUM CHANNEL.

```



```

      :: |||||:: ||||| :: |||||
7 IlecyProtyrGlyProProAspTyrTyrMetLysThrLeuGlnLeu1 23
482 .....TCTTCATATTGTTGATTC 464
23 eGlyLeuPheSerLeuProIleAsnPheLeuCysPheTyrPheValTyrP 40
463 TCCCTCCGCTCAATAAAAAAATAACACCGCATGCTTTTGTAT 414
:: |||||:: ||||| :: |||||
40 heTyrThrProLysGluSerLysPhe.....ArgTyrCysLeuAlaTyr 54
413 TTCGTTTTCGTTATCCATCTTCCTGTTGATTTCTCTGTC...CC 367
||| ||| :: |||||:: |||||
55 PheGlnPheIleAlaPheLeuValGlnValAspMetSerLeuValCysPr 71
366 GTCTCTCTCCCCCCCCCTCATATAAATATCTGTGTTAATTTGCTG 317
|||::|||:: |||||:: |||||
71 oGlyPhe.....TyrLeuPheProLeuMetGlyG 81
316 GGGTGGATMAATTTTGGGGGAAAAACTCATGTCCTCCCATCGGGGATAG 267
|| ::|||:: |||||:: |||||
81 LyTyrAsnLeuAlaGlnThrAsnArgLeuPheSerGlyHisGln..... 95
266 CTCGCACACAGTCACGTTGCTCTATCTCTGTTCAACACANTGGGACAC 217
|||::|||:: |||||:: |||||
96 .....ThrValValPheGlyTyrPheLysPheSerPheGluLe 108
216 ACAAGTTGCTCTCATGTTGGTTGACACAGTCATCGGACACTGCA.... 171
::|||:: |||||:: |||||
108 uProSerLeuLeuLeuLysPheIleTyrArgHisAsnAlaAlaAsnPr 125
171 ..... 171
125 heAsnProLysPheLysIleArgSerIleGlnTyrLeuValIleAla 141
170 ...TGCCACACAGTCCACGCACTGCTGT..... 144
|||::|||:: |||||:: |||||
142 ThrCysHisThrPheProPheValThrAlaIleCysLeuLysSerG1 158
143 .CTGTCATGCATGACTGTTGAGATTTATGATAGCTAGG..... 102
|||::|||:: |||||:: |||||
158 uLeuSerHisValGlnGlnValArgIleLeuGlnLysValArgArgTyrL 175
101 .....ACCAGATGCATAGACAGACTCACAGAAGATCCAGGCAGAA 60
|||::|||:: |||||:: |||||
175 LysLeuGlnThrArgGluThrAlaAsnTyrThrGlnAsnTyrProAsnCys 191
59 GCACACCTGGTG 48
|||::|||:: |||||:: |||||
192 LeuHisIleLeu 195

seq_name: sp_fungi:Q9P472

seq_documentation_block:
ID Q9P472 PRELIMINARY: PRT; 472 AA.
AC Q9P472;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CAMP-DEPENDENT PROTEIN KINASE PKAC CATALYTIC SUBUNIT.
GN PKAC.
OS Emerizella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emerizella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RT Shimizu K., Keller N.P.;
RT "Cloning of PKAC, a gene encoding a CAMP-dependent protein kinase
RT catalytic subunit of Aspergillus nidulans."
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1. SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF262987; AAF75762.1; -.

```

```

DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000961; Pkinase_C.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00069; Pkinase_1.
DR Pfam: PF00433; Pkinase_C_1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00220; S_TKc_1.
DR SMART: SM00133; S_TK_X_1.
DR PROSITE: PS00107; PROTEIN_KINASE_APP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR AMP-binding: kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 472 AA; 53233 MW; D3187698050A0D8A CRC64;

```

```

alignment_scores:
Quality: 77.00 Length: 139
Ratio: 1.305 Gaps: 7
Percent Similarity: 42.446 Percent Identity: 25.899

```

alignment_block:

US-09-823-101-3 x Q9P472 ..

Align seg 1/1 to: Q9P472 from: 1 to: 472

```

26 CTGCTGGGACCCATATGCCCTGTACACAGGTGCTTGGCCGTGATCTTCT 75
:: |||||:: ||||| :: |||||
312 IleHrThrPheLeuLysGlyThrProAspTyrLeuAlaProGluValVa 328
76 TGTGAGTCTCTGCTATGATCC.....TGGCTCTAGCGTATCC 113
::|||:: ||| ::|||
328 LAlaSerLysGlyTyrAsnLysSerValAspTyrPheSerLeuGlyLeu 345
114 TATATTCTCACAGCTCATATCGATGACACAGCAAGATGCTGGAGAC 163
|| |||||:: ::
345 euIlePheGluMetLeu..... 350
164 TGTGCGATGACGTTACAGATGACGTGTCACCAACAATGAGAGCAACT 213
|||::|||:: |||||
351 CysGly.....PheThr..... 354
214 TGTGTCACACANTGTTTGACAGAGT.....ATAGCACACAGTG 254
||| ||| ::|||
355 .....ProHrThrPaspGlnGlySerProValLysIleTyrGlnsni 369
255 ACNTGTGNGAGCTATCCCGATGGGGGACATGATTTTCCCCCAAAA 304
::|||:: |||
369 leuAlaGlnArgIleLys.....PheProProTyr 379
305 ATTATCCACCC.....CACCAATTAAATAACAAGATAT 339
::|||:: |||
380 LeuHisProAspAlaValAspLeuLeuSerArgLeuLeuThrSerAspLe 396
340 TTTATATAGGGGGGGGAGACAGACGGACAGCAAGAAATATACAGAC 389
:: |||||:: |||||
396 uThrLysArgLeuLysAsnLeuHisGlyLysProAspAspIleLysAsnH 413
390 GAAGATGATTAACGAA 406
|||::|||
413 IsProTyrPheAlaGln 418

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-4 to: A_Geneseq_1101.* out_format : pfs

Date: Jan 17, 2002 3:59 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-O=/cgn2_1/USPRO.spool/US09823101/runat_17012002_155137_22003/app_query.fasta_1.2239
-Db=A_Geneseq_1101 -OFMT=fastan -SUFFIX=n2p.rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext MINLEN=0 -MAXLEN=2000000000
-USER=US09823101_GCGN1_1195 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEXY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-4
Query length: 509
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 146.120000

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation	..	454	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB41897 +					101.50	186.21	0.0047	454	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB40851 +					99.00	170.65	0.0119	1310	1
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV22206 +					97.00	179.24	0.0144	361	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG23678 +					96.50	181.22	0.0152	266	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG23677 +					96.50	181.04	0.0152	271	1
/SID88/gcgdata/geneseq/geneseqp/AA1995.DAT:AAAR79912 +					96.50	171.60	0.0196	707	1
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAAB84052 +					96.50	171.60	0.0196	707	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB48964 +					96.00	171.60	0.0196	707	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB42438 +					96.00	169.62	0.0229	779	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAV68784 +					96.00	165.91	0.0253	1115	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAE04368 +					96.00	165.27	0.0253	1212	1
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV55954 +					96.00	165.10	0.0259	1233	1
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV55931 +					95.50	164.63	0.0281	1165	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB43016 +					94.00	166.06	0.0382	737	1
/SID88/gcgdata/geneseq/geneseqp/AA1994.DAT:AAAB56494 +					94.00	166.06	0.0382	737	1
/SID88/gcgdata/geneseq/geneseqp/AA1997.DAT:AAAM25019 +					94.00	166.06	0.0382	737	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAM39441 +					93.50	173.16	0.0351	323	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAG75411 +					93.50	172.45	0.0358	347	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAG75411 +					92.00	166.64	0.0541	374	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB43954 +					92.00	168.59	0.0542	376	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB39187 +					92.00	166.91	0.0567	446	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB67047 +					92.00	161.96	0.0647	737	1
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV29039 +					91.50	162.78	0.0702	611	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB36624 +					91.00	172.04	0.0609	215	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG15789 +					91.00	171.38	0.0619	230	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG52491 +					91.00	171.38	0.0619	230	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAV409410 +					89.50	159.98	0.0840	732	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG20196 +					89.50	170.83	0.0859	178	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49267 +					89.50	170.83	0.0859	178	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG20195 +					89.50	164.96	0.1005	323	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49266 +					89.50	164.96	0.1005	323	1
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49265 +					89.50	164.69	0.1042	332	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB93238 +					89.00	155.03	0.1454	798	1
/SID88/gcgdata/geneseq/geneseqp/AA1998.DAT:AAAB53571 +					89.00	153.04	0.1533	976	1
/SID88/gcgdata/geneseq/geneseqp/AA1998.DAT:AAAB41377 +					89.00	153.04	0.1533	976	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAM40294 +					88.00	151.43	0.1600	1150	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAU03590 +					88.50	161.00	0.1375	392	1
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB92959 +					88.50	161.00	0.1375	392	1

/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB93038 +	88.50	161.00	0.1375	392
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB94332 +	88.50	161.00	0.1375	392
/SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV40999 +	88.50	156.73	0.1542	605
/SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB18188 +	88.50	156.08	0.1569	646
/SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB99675 +	88.50	149.61	0.1864	1246
seq_name: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB41897				
seq_documentation_block:				
ID	AAAB41897	standard:	Protein:	454 AA.
XX				
AC	AAAB41897:			
XX				
DT	08-FEB-2001 (first entry)			
XX				
DE	Human ORFX ORF1661 polypeptide sequence SEQ ID NO:3322.			
XX				
KM	Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;			
KM	vulnerary; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;			
KM	anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;			
KM	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;			
KM	hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;			
KM	antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;			
KM	antianemic; gene therapy; cancer; proliferative disorder; hypertension;			
KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;			
KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;			
KM	cholesterol ester storage; systemic lupus erythematosus; infection;			
KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;			
KM	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;			
KM	bone damage; cartilage damage; antineoplastic disease; coagulation;			
KM	thrombosis; contraceptive.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200058473-A2.			
XX				
PD	05-OCT-2000.			
XX				
PE	31-MAR-2000; 2000WO-US08621.			
XX				
PR	31-MAR-1999; 98US-0127607.			
PR	02-APR-1999; 98US-0127636.			
PR	05-APR-1999; 99US-0127728.			
XX	30-MAR-2000; 2000US-0540763.			
PA	(CUBA-) CUBAGEN CORP.			
XX				
PI	Shinkets RA, Leach M;			
XX				
DR	WPI; 2000-602362/57.			
XX				
DR	N-PSDB; AAC76106.			
XX				
PT	Novel nucleic acids and peptides derived from open reading frame X,			
PT	useful for treating e.g. cancers, proliferative disorders,			
XX	neurodegenerative disorders and cardiovascular disease -			
PS	Claim 11; Page 2511-2512; 5507pp; English.			
XX				
CC	AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,			
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX			
CC	sequences have activities such as: cytosolic; hepatotropic; vulnerary;			
CC	antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;			
CC	osteopathic; anticonvulsant; antiallergic; immunosuppressant;			
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;			
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;			
CC	antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;			
CC	antianemic; gene therapy; cancer; proliferative disorder; hypertension;			
CC	the presence of or predisposition to, or preventing or treating			
CC	pathological conditions associated with an ORFX-associated disorder. The			
CC	nucleic acids can be used to express ORFX proteins in gene therapy			
CC	vectors. The proteins and nucleic acids may be used to treat cancers,			
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,			
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,			

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 454 AA;

Alignment_scores:
Quality: 101.50 Length: 162
Ratio: 0.976 Gaps: 5
Percent Similarity: 64.198 Percent Identity: 24.691

Alignment_block:

US-09-823-101-4 x AAM41897 ..

Align seg 1/1 to: AAM41897 from: 1 to: 454

30 ATCAGCGAAGAAACGTAATCTATCA.....GAGCAAAAAAANA 70
262 ValArgGluAlaGlnSerPheIleSerAlaAlaIleGluProGlnSerI 278
71 AAAAAAANAACCCGGGGGAAACGAGGACATAGCAGTG.. 118
278 yLsSerAsnGluArgLysGlyIArgSerArgSerHisThrArgSerI 295
119TCCGGGCTGGAACCTGTGTACCCGGGTACAAAC 154
295 ySerArgSerSerLysSerHisSerArgArgLysArgSerGlnSer 311
155 TCCAAAGAGAAACATGTCACACGCTCCACCTACCGCGACACAGCTA 204
312 LysHisArgSerArgSerHisArgSerArgSerArgGlnLysAspArg 328
205 GAGAAACATCTACTAGAACTCAACATATATATACATAGTACAGAGAA 254
328 gArgArgSerLysSerProHisLysLysArgSerLysSerArgGluArg 345
255 ACGCACATAGAGCTACCGCATTAAGTACAACTCGCAAGACACGCTAT 304
345 rGlySerArgSerArgSerHisSerArgAspLysArgLysAspThrArg 361
305 CAGAGACTGACGAGAGAGAAACGACGACAGAGAGACAAACACAGAGC 354
362 GluArgLysLysGluArgValLysGluLysAspArgLys. 377
355 CAGCGTACGCGATGAGGACGAGAGCGAGATACCGGAAAGAAACGGGAA 404
378 GluArgGluArg...GluLysGluArgGluLysGluLysGluArgGly 393
405 AAGACAGATGAAGTATCTCAGTTGACGAGAACCAAAAGCAGAGAGC 454
393 sAsnLysAspArgAspLysGluArgGluLysAspArgGluLysAspLysG 410
455 ATGTGTATCAAGAAACGACGCGACAGACAGC 486
410 LuLysAspArgGluArgGluArgGluLysGlu 420

seq_name: /SID58/gcgcdata/geneseq/geneseq/AA2001.DAT:AAM40851

seq_documentation_block:

ID AAM40851 standard; Protein; 1310 AA.

XX AAM40851;

XX 22-OCT-2001 (First entry)

XX Human polypeptide SEQ ID NO 5782.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60007.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 5782; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1310 AA;

Alignment_scores:
Quality: 99.00 Length: 158
Ratio: 1.138 Gaps: 5
Percent Similarity: 55.063 Percent Identity: 25.949

Alignment_block:

US-09-823-101-4 x AAM40851 ..

Align seg 1/1 to: AAM40851 from: 1 to: 1310

28 TCATCAGGAGAAACGTAAT.....TCTATCAGAGAGAA 62

666 SerSerGlyGluLysGlyGlyTyrGluIlePheValLysAspThr 682

63 AAAAAAANAACCCGGGGGAAACGAGGACATAG 112

682 rHisGluLysSerLysAsnLysAsnArgAspLys..... 694

```
113 CGAGTGTCCCGGGCTGGAACCTGTGTACCCGCTACAAACTCCAAAG 162
695 ..... GlyIuLysGluLysLys 700
163 AGGAACATTTTCACACGCGCTCCACTACCGGCACACAGCTGAGAAACA 212
701 ArgAspSerSerLeuArgSerArgSerLysArgSerLysSerSerLeuI 717
213 CATCTACTAGAACTCACACATATATACATAGTACAGAGAAAGCCGACA 262
717 slySerArgLysLeuIhSerGluSerArgSerArgAlaArgLysArg 733
263 TAGAGCTACTCCGATAAAGTACAACTGCAGAAAGACCGCTATCAGAGACT 312
734 ..... SerSerLysSerLysSerLysArgSerGlnThr 744
313 GACGAGACA..AGAAACGACACAGAGAGACAAACACAGACGCCAGCG 359
745 ArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 761
360 TGAGCGCCGATGAGCGACGCGAGATACCCGAAAGAAAGCGGAAAGAC 409
761 gserLysSerArgGlyArgArgSerValSerLysGluLysArgLysArgS 778
410 AGATGAAGTAATCTCAGTTGACGAAAGACCAAAAGCGAGAGACGATGTT 459
778 erProLysHisArgSerLysSerArgGluArgLysArgLysArgSer... 793
460 GATCAAGAACACCGCGCGAGACA 483
794 SerSerArgAspAsnArgLysThr 801

seq_name: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:AAV22206
seq_documentation_block:
ID   AAV22206 standard; Protein: 361 AA.
AC   AAV22206;
XX
DT   13-SEP-1999 (first entry)
DE   Biorhythm marker protein.
KW   Biorhythm marker gene; circadian rhythm.
XX
OS   Rattus sp.
PN   JP11169185-A.
PD   29-JUN-1999.
PE   12-DEC-1997; 97JP-0362890.
PR   12-DEC-1997; 97JP-0362890.
XX
PA   (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
DR   WPI: 1999-422623/36.
DR   N-PSDB: AAX84569.
XX
PT   Structure and function of a new biorhythm marker gene - and protein
XX   encoded by it
XX
PS   Claim 7; Page 14-16; 23pp; Japanese.
XX
CC   This sequence is the biorhythm marker protein of the invention. The
XX   protein is also referred to as a circadian rhythm protein..
SQ   Sequence 361 AA;
```

```
alignment_scores:
  Quality: 97.00      Length: 158
  Ratio: 1.032      Gaps: 5
  Percent Similarity: 59.494      Percent Identity: 22.152

alignment_block:
US-09-823-101-4 x AAV22206 ..
Align seg 1/1 to: AAV22206 from: 1 to: 361

48 TTCTATCAAGAGCAAAAAA.....AAACCG 91
111 .....
47 PheLeuProGluSerArgLysSerAsnArgSerArgSerArgGluArgSerAr 63
92 GCGGAAACCGAGGACATAGCGAGTCCCGGCTGAAACTGTGCTAC 141
111 .....
63 gserHisThrArgSerLysSerArg..... 71
142 CCGGCTACAAACTCCAAAGAGAAATTTGTCAACGCGTCCACTACC 191
72 .....SerSerLysSerHisSerArgArgLysArgSerGlnSer 85
192 GCGCACACAGCTAGAAACATCTACTAGAAACTCACACATATATATAC 241
111 .....
86 LysHisArgSerArgSerHisAsnArgSerArgSerArgGluLysAspAr 102
242 ATAGTACAGAAACGCCACAT.....AGAGTACTCGCATTA 279
102 gaArgArgSerLysSerProHisLysArgSerLysSerArgGluArgA 119
280 AGTACAGTCCGCAAGACACGATATCAGAGACTGACAGAGAAAGCA 329
119 rglYSerArgSerArgSerArgSerArgAspLysArgLysAspThrArg 135
330 CGACAGAGAGAACCAACAGACGCGGAGCGCGCAT..... 369
136 GluLysValLysGluArg..ValLysGluArgGluArgLysGluLysG 152
370 ....GAGCGACAGCGCAGATACCCGAAAGAAACGCGGAAAGACAGATG 414
152 luArgGluLysGluArgGluLysAspLysGluArgGlyLysAsnLysAsp 168
415 AAAGTAATCTCAGTTGACGAGAACCAAAAGCAGAGACGATGTGATCA 464
169 LysAspArgGluLysGluLysAspHisGluLysGluArgAspLysGluL 185
465 AAGAAACCGACCGCGAGACAGAC 486
185 sgluLysGluGlnAspLysAsp 192

seq_name: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:AA23678
seq_documentation_block:
ID   AA23678 standard; Protein: 266 AA.
XX
AC   AA23678;
XX
DT   17-OCT-2000 (first entry)
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27074.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PP   25-FEB-2000; 2000EP-0301439.
XX
```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137552.
PR 04-JUN-1999; 99US-0137552.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139466.
PR 18-JUN-1999; 99US-0139467.
PR 18-JUN-1999; 99US-0139468.
PR 18-JUN-1999; 99US-0139469.
PR 18-JUN-1999; 99US-0139470.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147203.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 96.50      Length: 128
  Ratio: 1.664      Gaps: 3
Percent Similarity: 45.312      Percent Identity: 22.656
```

alignment_block:
US-09-823-101-4 x AAG23678 ..

Align seg 1/1 to: AAG23678 from: 1 to: 266

```
28 TCATCGGGGAAAAACGTATCTATCAAGAGAAAAA
|||||
56 SerGlyLysLysLysArgTnTArgArgLysArgLysLys
78 AAAAAAAACCCGGGGGGAACCAAGGGGACATAGCGAGTCCGGGCT
|||||
72 sLysLysLysLysGlyGlyArgAspCysArgIleCysHisLeuPro...
128 GGAACCTGTGTACCCGGCTACAAACCTCCAAAGAGAAACATTGTCAA
|||||
178 CGGGCTCCACTACCGCGCACACAGCTAGAGAAACACATCTACTAGAACT
95
228 CACACATATATACATAGTACAGAGAAACGCACATAGAGCTACGCAAT
95
278 AAGGTCACTGCGCAAGACAGCTATCAGAGACTCAGCGAGAGAGAAC
96
328 GAGCAGACGAGAGACAACCAAGCAGCGACGTCGAGCGGCGATGAGCGAGA
104 AspSerAspAspAspGluAspGluAspGluAspGluGluGluGluGluGlu
```

```
378 GCCGACATAC.....CCGAAAGAACGGCG 402
|||||
120 uGluGluTyrTyrGlyLeuProLeuGlnLeuGly 131

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG23677
seq_documentation_block:
ID AAG23677 standard; Protein; 271 AA.
AC AAG23677;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27073.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0122180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139753.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149725.
PR 20-AUG-1999; 99US-0149726.
PR 20-AUG-1999; 99US-0149727.
PR 20-AUG-1999; 99US-0149728.
PR 20-AUG-1999; 99US-0149729.
PR 20-AUG-1999; 99US-0149730.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 96.50 Length: 128
Ratio: 1.664 Gaps: 3
Percent Similarity: 45.312 Percent Identity: 22.656

alignment_block:

US-09-823-101-4 x AAG23677

Align seg 1/1 to: AAG23677 from: 1 to: 271

28 TCATCAGGGAACGTAATTCATCAAGAGCAAAAAAAAAAAAAAAAAAAAA 77

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AY55954

seq_documentation_block:

ID AAY55954 standard; Protein; 1233 AA.

AC AAY55954;

DT 18-FEB-2000 (first entry)

DE Mouse STE20-related protein kinase NIK_m.

KM Antihypertatic; antiatherobic; antinflammatory; antiallergic; osteopathic;
KM antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
KM neuroprotective; cardiant; cerebroprotective; cytostatic; antidiabetic;
KM valnery; STE20, protein kinase; SLK2; SLK3; SLK4; SLK5, SLK6, SLK7;
KM ZC1, ZC2, ZC3, ZC4, KHS2, SOLTU, SULT3, GSK3, PAK1, PAK5; antagonist;
KM antibody; gene therapy; rheumatoid arthritis; arteriosclerosis; asthma;
KM inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
KM rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
KM myocardial infarction; cardiovascular disease; stroke; renal failure;
KM oxidative stress-related neurodegenerative disorder; Parkinson's disease;
KM amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
KM ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
KM meningeal disorder; growth regulation; wound healing; T cell activation;
KM immunosuppressant.

OS Mus sp.

PN W09953036-A2.

PD 21-OCT-1999.

PF 13-APR-1999; 99WO-US08150.

PR 14-APR-1998; 98US-0081784.

PA (SUGEN-) SUGEN INC.

PI Plowman G, Martinez R, Whyte D

WPI: 1999-611301/52.

PT Novel kinase-related polypeptides used for the diagnosis and treatment of kinase-related diseases and disorders -

PS Disclosure; Page 339-343; 387pp; English.

CC This sequence represents a novel SRE20-related protein kinase. The
CC invention relates to nucleic acid molecule encoding a kinase polypeptide
CC selected from S1LW2, S1LW3, S1LW4, S1LW5, S1LW6, S1LW7, ZC1, ZC2, ZC3,
CC ZC4, KHS2, S1U01, S1U03, GEX2, PAK4 and PAK5. The proteins are used to
CC identify agonists and antagonists, and to raise antibodies. The
CC polypeptides are useful in gene therapy protocols. The polynucleotides/
CC polypeptides, antibodies, antagonists and agonists may be used to treat
CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.
CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
CC rheitis, autoimmunity, and organ transplantation, chronic inflammatory
CC pelvic disease, multiple sclerosis, organ transplantation, myocardial
CC infarction, cardiovascular disease, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,
CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
CC mellitus, fibrotic and mesangial disorders. The proteins may also be
CC useful for cell growth regulation (e.g. in wound healing), T cell
CC activation, mitosis control, and as immunosuppressants.

SQ Sequence 1233 AA;

alignment_scores:

Quality: 95.00

Ratio: 1.043

Length: 174

Gaps: 5

Percent Similarity: 52.874 Percent Identity: 21.8339

alignment_block:

US-09-823-101-4 x AAY55954

Align seg 1/1 to: AAY55954 from: 1 to: 1233

43 CGTATTCTATCAAGGAAAAAAAAAAAAAAAAAAAAA..... 84
302 LysaspHstIleasphgrrtArgrLysLysArgLyGIuLysAspGIuTh 318
85AAAC 88
318 rGLuTyGLuTySerCIySerGIuGIuGIuGIuGIuValProGIuG 335
89 CCGGGGGGAACCAAGGAGCATAGCCAGTGTCGCCGGCTGGAAA...CTG 135
335 InGIuGIuGIuProSerSerLevalAsnValProGIyGIuSerThrLeu 351
136 TGTACCCTGCCTCAAAACTCATAAAGAGAACAATGTTCACAGCGCTCC 185
352 ArgArGrSpHeLeuArGrLeuGIuGIuGIuGIuGIuGIuGIuGIuGIu 368
186 ACTACCGCCACACACGTAGAGAAACACATCTACTAGAAATCACACAT 235
368 aleuArGrArgGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 379
236 ATATACATAGTACAGAGAAAGCCACATAGAGCTACTGCA..... 276
379 InLeuArGrGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 395
277TAAGTACAACTGGCGCAAAAGACAGTACAGAGATGA 314
396 LysArGrILeGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 412
315 CGAAGAGAGAAAGACGACAGAGAGAACAAACACAGACCGCGTAGAC 364
412 nArGrArGIuArGrGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 429
365 GCGATGTAGCGACGCGACATACCCCAAAAGAAAGCGGGAAGACAGATG 414
429 rGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 445
415 AAAGTAATCTCAGTTGACGAGACCAAAGCCGGAAGACAGATTTATCA 464
446 GIu.....GIuGIuGIuArGrArGrAlaGIuGIuGIuGIuLysAr 458
465 AAGAAACGAGCGGAGACAGAC 486
458 gArGrValGIuArGrGIuGIuGIu 465

seq_name: /SIDS8/gc9data/geneseq/geneseq/PAA199.DAT:AAV55931

seq_documentation_block:
ID AAV55931 standard; Protein: 1239 AA.
XX AC AAV55931:
XX DT
XX DE
XX Human ZC1 protein.

Antirheumatic; antiarrhythmic; antinflammatory; antiallergic; osteopathic;
antipsoriatic; antiarteriosclerotic; antiaesthetic; immunosuppressive;
neuroprotective; cardiact; cerebroprotective; cytotactic; antidiabetic;
vulnerly; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
ZC1, ZC2, ZC3, ZC4, KHS2, SULU1, SULU3, GERK3, PAK4; PAK5; antagonist;
antibody; gene therapy; rheumatoid arthritis; artherosclerosis; asthma;
inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
myocardial infarction; cardiovascular disease; stroke; renal failure;
oxidative stress-related neurodegenerative disorder; Parkinson's disease;
amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;


```
67 AAAAAAAAAAAAAAAAACCCGGGAACCA..... 102
   ::|||::||| |||::||| |||
468 GlnLysAsnLysLeuLysGlnProGlyGlySerPheArgAspAspAs 484
103 .....GGGACATAGCGAGTGTCCGGCTGGAACACTGTGTACCCGG 145
   ::|||::||| ||| ||| |||
484 pAspIleAsnSpValAlaSerMetLagIyValAsnLeuSerGluIuS 501
146 CTACAAACTC.....CAAAAGAGCAA 168
   ::|||::||| ||| ||| |||
501 eAlaIArgIleuAlaThrAsnSerGluLeuValGlyThrLeuThrArg 517
169 CATTGTCAA.....CGCGCTCCACTACCGCGCACACA 200
   |||::||| ||| ||| |||
518 SerCysLysAspGluThrPheLeuLeuGlnAlaProLeuGlnArgI 534
201 GCTAGAG.....AAACAC.....ATTACTAGAAACTCA 229
   ||| ||| ||| ||| |||
534 eLeuGluIleGlyLysLysHisGlyIleThrGluLeuHisProAspVal 551
230 CAACATATATA...CATAGTACAGAAAGCCACATAGAGTACTCGCA 276
   ::|||::||| ||| ||| ||| |||
551 aISerTYrValSerHisAlaThrGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTCAACTCGCGCAAGACACCTATCAGACTGACGAGAGAAGAA 326
   ::|||::||| ||| ||| ||| |||
567 .....GluLysIleSerGluThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGACAAACCAACGAGCCAGCGT. 360
   ||| ||| ||| ||| ||| |||
577 nPheSerTYrLysAspAspArgTYrGlnGlnAlaSerAspValArgA 594
361 .....GAGCGC 366
594 IagIleuLysPhePheGlnLeuAspGlnIleGlnLysGlnArgLys 610
367 GATGAGCGAGAGCGAGATACCGGAAGAAAGGGAAGACAGATGAA 416
   ||| ||| ||| ||| ||| |||
611 AspGlnGlnLysArg.....GluIleLeuMetArgAlaAlaLys 623
417 AGTAATCTCAGTTCAGCAGAGAC..... 438
   ||| ||| ||| ||| |||
623 sSerArgSerArgGlnGlnAspProGlnGlnLeuArgLeuLysGlnLysA 640
439 .....CAAAAGCGAGAGACGAT... 456
640 IalysGlnMetGlnGlnGlnGlnLeuAlaGlnMetArgGlnArgAspAla 656
457 .....GTTGATCAAGAAGAACAGCGCGAGACAGAGA 485
   ::|||::||| ||| ||| ||| |||
657 AsnLeuThrAlaLeuAlaIleGlyProArgLysLysArgLysValAs 673
486 CTGT.....GCAGAAGTGG 501
   ||| ||| ||| ||| ||| |||
673 pCysProGlyProGlySerGlyAlaGlnGlySer 684
```



```

146 CTACAAACTC.....CAAAAGAGAAA 168
      :::::::::::
501 erAlaIarglleuAlaIthrAsnsergluLeuValglYthrleuThrArg 517
169 CATTTGCAA.....CGCGCTCAGTACCGCGACACA 200
      :::::::::::
518 SerCysLysAspGluThrPheLeuLeuGlnAlaProLeuGlnArgArgI 534
201 GCTAGAG.....AAACAC.....ATCTACTAGAAACTCA 229
      :::::::::::
534 eleuGluIleGlyLysShsIglYIleThrGluLeuHisProAspValY 551
230 CAACATATATA...CATAGTACAGAGAAAGCCACATAGAGCTACTCGCA 276
      :::::::::::
551 alSerTyValSerHisAlaThrGlnGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTACAACTGCGCAAGACAGCTATCAGAGACTGACGAGACAGAGAAA 326
      :::::::::::
567 .....GluLysIleSerGluThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGAGACAAACCAACGACGCGCCT. 360
      :::::::::::
577 nPheSerTyLysAspAspArgTyrgIuGlnAlaSerAspValArgA 594
361 .....GAGCGC 366
594 IaGlnLeuLysPhePheGlnGlnLeuAspGlnIleGluLysGlnArgLys 610
367 GATGAGCAGAGAGCGCAGATACCCGAAAGAGGGGAAAGACAGATGAA 416
      :::::::::::
611 AspGlnGlnGlnArg.....GluIleLeuMetArgAlaAlaIly 623
417 AGTAATCTCAGTTGACGAAAGC..... 438
      :::::::::::
623 sSerArgSerArgGlnGlnAspProGluGlnLeuArgLeuLysGlnLysA 640
439 .....CAAAAGCAGAGAGACGAT... 456
      :::::::::::
640 IaLysGluMetGlnGlnGlnIuLeuAlaGlnMetArgGlnArgAspAla 656
457 .....GTTGATCAAGAGAAACGAGCGGAGACAGAGA 485
      :::::::::::
657 AsnLeuThrAlaLeuAlaIleGlyProArgLysArgLysValAs 673
486 CTGT.....GCAGAAAGCTGCG 501
      :::::::::::
673 pCysProGlyProGlySerGlyAlaGluGlySer 684

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep.us-08-646-715-16

seq_documentation_block:
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAIFS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-16

alignment_scores:
      Quality: 94.00      Length: 228
      Ratio: 0.969      Gaps: 13
Percent Similarity: 42.544      Percent Identity: 23.246

alignment_block:
US-09-823-101-4 x US-08-646-715-16 ..

Align seg 1/1 to: US-08-646-715-16 from: 1 to: 737

67 AAAAAAAAAAAAAAAAAAACCAGCGGGAACCA..... 102
      :::::::::::
468 GlnYsaSnLysLeuLysGluProGlyGlySerPheArgAspAspAs 484
103 .....GGGACATAGCAGAGTGTCCGCGTGAAGTGTGTACCCGG 145
      :::::::::::
484 pAspIleAsnAspValAlaSerMetAlaGlyValAsnLeuSerGluG 501
146 CTACAAACTC.....CAAAAGAGAAA 168
      :::::::::::
501 erAlaIarglleuAlaIthrAsnsergluLeuValglYthrleuThrArg 517
169 CATTTGCAA.....CGCGCTCAGTACCGCGACACA 200
      :::::::::::
518 SerCysLysAspGluThrPheLeuLeuGlnAlaProLeuGlnArgArgI 534
201 GCTAGAG.....AAACAC.....ATCTACTAGAAACTCA 229
      :::::::::::
534 eleuGluIleGlyLysShsIglYIleThrGluLeuHisProAspValY 551
230 CAACATATATA...CATAGTACAGAGAAAGCCACATAGAGCTACTCGCA 276
      :::::::::::
551 alSerTyValSerHisAlaThrGlnGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTACAACTGCGCAAGACAGCTATCAGAGACTGACGAGACAGAGAAA 326
      :::::::::::
567 .....GluLysIleSerGluThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGAGACAAACCAACGACGCGCCT. 360
      :::::::::::
577 nPheSerTyLysAspAspArgTyrgIuGlnAlaSerAspValArgA 594
361 .....GAGCGC 366
      :::::::::::

```

```

594 l a g l i n l e u l y s p h e p h e g l i n g l i n l e u a s p c l i l l e g l u l y s g l n a r g l y s 610
367 G A T G A G G C A G A C G C A G A T A C C G A A A A C G G G A A A A G A C A G A T G A A 416
||||| ||||| ||||| |||||
611 A s p g l u g l i n l u a r g ..... G l u l l e u m e t a r g l a l a l a l y 623
417 A C T A A T C T C A G T T G A C G A G A C ..... 438
||| ::|||
623 s s e r a r g s e r a r g l i n g l u a s p p r o g l u i n l e u a r g l e u l y s g l n l y s a 640
439 ..... C A A A C G C A G A A A C A G C A T ... 456
||||| ||||| |||||
640 l a l y s g l u m e t c i n g l i n g l i n l e u a l a l a t h r p r o g l i n a r g s p a l a 656
457 ..... G T T G A T C A A A G A A A C C A G C G C G A C A G A 485
||||| ||||| |||||
657 A s n l e u t h r l a l e u a l a l a l e g l y p r o a r g l y l y s a r g l y s v a l a s 673
486 C T G T ..... G C A G A A G G T G C G 501
||||| ||||| |||||
673 p c y s p r o g l y p r o g l y s e r g l y a l a g l i n l y s e r 684

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-990-114-3

seq_documentation_block:

```

: Sequence 3, Application US/08990114
: Patent No. 5932475
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Yue, Henry
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/990.114
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0451 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 714 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 128842
: US-08-990-114-3

```

```

alignment_scores:
  Quality: 91.50      Length: 170
  Ratio: 0.943      Gaps: 7
  Percent Similarity: 57.059      Percent Identity: 24.118

```

alignment_block:

US-09-823-101-4 x US-08-990-114-3 ..

Align seg 1/1 to: US-08-990-114-3 from: 1 to: 714

```

38 A A A A C G T A A T C T A T C A A G A G A A A A A ..... A A 69
||||| ||||| ||||| |||||
61 L y s l y s v a l v a l v a l s e r g i n t h r l y s l y s v a l a l a v a l p r o t h p r o a l 77
70 A A A A A A A A A A A A A A A C C G G G G A A A C C A G G G A C A T A G C A G T G T 119
||||| ||||| ||||| |||||
77 a l y s l y s a l a l a v a l t h r p r o g l y l y s l y s a l a l a l a t h r p r o a l 94
120 C C C G G C T G G A A A C T G T G T A C C G C T A C A A A A C T C A ... A A A G A G A 166
||||| ||||| ||||| |||||
94 y a l y s a l a . v a l t h r p r o a l a l y s a l a v a l a l a t h r p r o g l y l y s g l 110
167 A A C A T T G T C A A G C G C C T C A C T A C G C G C A C A C A G C T A G A A A C A C A T C 216
||||| ||||| ||||| |||||
110 y a l a t h r g l i n a l a l y s a l a l e u v a l a l a t h r p r o g l y l y s g l y a l a l 127
217 T A C T A G A A A C T C A A C A T A T A T A C A T A G T A G A G A A A C G C A C A T A G A 266
||| ||||| ||||| |||||
127 a l ..... T h r p r o a l l y s 131
267 G C T A C T C G C A T A A G T A C A C T G C C A A A G A C A G T A T A G A G A C T G A C G 316
||||| ||||| ||||| |||||
132 G l y a l a l a y s a n g l y l y s a s n . A l a l y s l y s g l u a s p s e r a s p g l u a s p g 148
317 A G A A G A A A A C A G C A G C A G A G A G A C A A A C A A C G A C G G T G A G C C C 366
||| ||||| ||||| |||||
148 l u a s p a s p a s p a s p g l u a s p a s p s e r a s p g l u a s p g l u a s p g l u 164
367 G A T G A G G C A G A C G C A G A T A C C G A A A A G A A C G G G A A A A G A C A G A T G A A 416
||||| ||||| ||||| |||||
165 G l u g l u a s p c l u p h e g l u p r o p r o v a l v a l l y s g l y l y s ... G l n g l y l y 180
417 A G T A ..... A T C T A G T T G A C G A A G A C C A A A A C G A A G 451
||||| ||||| ||||| |||||
180 s v a l a l a l a l a l a l a p r o l a s e r g l u a s p g l u a s p g l u g l u a s p g 197
452 A C G A T T T G A T C A A A A A C C A G C G G A C A G A C A C T G T C A G A A G T G G G 501
||||| ||||| ||||| |||||
197 l u g l u g l u g l u g l u a s p g l u g l u g l u g l u a s p a s p e r g l u g l u 213
502 A G G C G 507
||||| ||||| ||||| |||||
214 G l u a l a 215

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-370-368-9

seq_documentation_block:

```

: Sequence 9, Application US/09370368
: Patent No. 6258932
: GENERAL INFORMATION:
: APPLICANT: Anders Valhne
: TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
: TITLE OF INVENTION: AND METHODS OF USE THEREOF
: FILE REFERENCE: TRIPEP.003A
: CURRENT APPLICATION NUMBER: US/09/370.368
: CURRENT FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 9
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Moloney Murine Leukemia Virus
: US-09-370-368-9

```

EARLIER FILING DATE: 1999-03-29

APPLICANT: Hillman, Jennifer
APPLICANT: Corlew, Neil C

```
APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,801
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0354 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNON01
CLONE: 2272281
US-08-903-801-1

alignment_scores:
Quality: 86.50 Length: 143
Ratio: 1.095 Gaps: 5
Percent Similarity: 55.245 Percent Identity: 25.175

alignment_block:
US-09-823-101-4 x US-08-903-801-1 ..

Align seg 1/1 to: US-08-903-801-1 from: 1 to: 290

54 CAAGAGGAAAAAAAAAAAAAAAAAACCAGGCGGAAACCG 103
   :::::::::::::::::::: :::::::::::
149 LysLysGluArgLysLysArgAspGluLysAlaLysLeuAr 165
   :::::::::::::::::::: :::::::::::
104 GGGACATAGCGGTGCCGGCTGGAACGTGTGACCGGCTACA 153
   :::::::::::::::::::: :::::::::::
165 GlySerLeuLysGluLysGluLysLeuGluThr.....GlyLysAs 179
   :::::::::::::::::::: :::::::::::
154 CTCCAAGAGGAACATTGTCACGCGCTCCACTACCGCGCAC 203
   :::::::::::::::::::: :::::::::::
179 spGlnSerLysGlnLysGlnSerGlnArgLysPheGluLys 195
   :::::::::::::::::::: :::::::::::
204 AGAGAAACATCTACTAGAAACTCACAACATATATACATA 253
   :::::::::::::::::::: :::::::::::
196 LysSerLysValThrValAspThrGlyValIleProAla 212
   :::::::::::::::::::: :::::::::::
254 AACGCGACATAGAGCTACTCGCATTAAGTACAACTCCG 303
   :::::::::::::::::::: :::::::::::
212 Ys.....AlaGluThrProThrAlaLacIuspsp 222
   :::::::::::::::::::: :::::::::::
304 TCAGAGACTGACGAGAGAAAGACGACGAGAGACAAACG 353
```

```
223 AsnGluGlyAspLysLysLys.....AspLysLysLys 235
   :::::::::::::::::::: :::::::::::
354 CCAGCGTGAGCGCGATGAGCGAGCGAGATACCGAAAGAG 403
   :::::::::::::::::::: :::::::::::
235 LysSerLysGluLysGluLysGluLysLysLysLysLys 252
   :::::::::::::::::::: :::::::::::
404 AAAGACAGTGAAGTAAATC.....TCAGTT 429
   :::::::::::::::::::: :::::::::::
252 YsaLThrValLysAlaMetGlnGluAlaLeuAlaLysLeu 268
   :::::::::::::::::::: :::::::::::
430 GACGAGACCAAAACGAGAGACGAT 456
   :::::::::::::::::::: :::::::::::
269 GluGluArgGlnLysArgLysGluLys 277
   :::::::::::::::::::: :::::::::::

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-295-055-1

seq_documentation_block:
; Sequence 1, Application US/09295055
; Patent No. 6232440
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/295,055
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/903,801
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0354 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNON01
; CLONE: 2272281
; US-09-295-055-1

alignment_scores:
Quality: 86.50 Length: 143
Ratio: 1.095 Gaps: 5
Percent Similarity: 55.245 Percent Identity: 25.175

alignment_block:
US-09-823-101-4 x US-09-295-055-1 ..
```

TELEFAX: (510) 222-9758

OTHER INFORMATION: /note= "Vicillin from G. max"

FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0465 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HEARNOT01
CLONE: 307624
US-09-016-000-2

alignment_scores:
Quality: 81.00 Length: 164
Ratio: 1.025 Gaps: 7
Percent Similarity: 48.171 Percent Identity: 25.000

alignment_block:
US-09-823-101-4 x US-09-016-000-2 ..

Align seg 1/1 to: US-09-016-000-2 from: 1 to: 688

```
19 ATGTGATATCATCAGGAAAAAAGCTATTCATATCAGGAAAAA 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 MetCysValAspAspAlaTyrValArgArgMetAlaAlaGluAlaThrGlu 238
69 AAAAAAAGAAAAAAGAAAAAAGCGGGGGAAGCCAGGAGCATGCGAGTG 118
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 utrpglnlvsalaglyalaproprosererglyseralalavaserthra 255
119 TCCCGGGCTGAAACTGTGTACCGGCTACAAAGCTCCAAAGAGAGAA 168
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 larpoglnlnlvs.....ProIleGlyIleSerLysAsnLys 268
169 CATGTCAAGCGCTCCACTACCGGCACACAGCTAGAGAAACACATCTA 218
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 .....LysLysLysLeuLysLys..... 274
219 CTGAAGACTCACACATATATACATAGTACGAGAAAGCCACATAGAGC 268
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 ...LysGlnLysThrGlnAlaGluLeuLeuGlnLysArgLeuGlnLys 290
269 TACTCGCATTAAGTACACTCGCAAGAGACAGCTATCAGACTGAC... 315
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 leGluGluLeuGlnArgGluAlaGlnArgLysIleIleGlnGluAsnIle 306
316 .....GAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 ThrSerAlaIleProSerAsnAspGlnAspGlyLysIleTyrCysProGlu 323
343 .....AAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 385
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 LysLeuLysThrThrGlyLeuGlnGluAlaAlaGlnAlaGlnLysThr... 338
386 ACCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ...AlaLysAspAsnGlyLysAlaGlnArgLysIleGlnLysGlnAspAla 354
436 GAG.....CAAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

355 GluLysGluAsnIleGluLysAspGluAspAlaAspGln 368

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-245-041-11

seq_documentation_block:

; Sequence 11, Application US/09245041

; Patent No. 6274339

; GENERAL INFORMATION:

; APPLICANT: Moore, K.

; APPLICANT: Nagle, D.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT

; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY

; FILE REFERENCE: 7853-136

; CURRENT APPLICATION NUMBER: US/09/245,041

; EARLIER FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/093,630

; EARLIER FILING DATE: 1998-07-21

; EARLIER APPLICATION NUMBER: 60/104,978

; NUMBER OF SEQ ID NOS: 131

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 11

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: SITE

; LOCATION: all Xaa positions

; OTHER INFORMATION: Xaa-unknown amino acid

US-09-245-041-11

alignment_scores:
Quality: 80.50 Length: 44
Ratio: 2.516 Gaps: 2
Percent Similarity: 72.727 Percent Identity: 47.727

alignment_block:
US-09-823-101-4 x US-09-245-041-11 ..

Align seg 1/1 to: US-09-245-041-11 from: 1 to: 351

```
37 AAAAAAGTATTCATCAAGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 86
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 45
87 ACCCGGGGGAAGCAAGCGGAGCATAGAGAGTGTCCGCGGCTGGAAGTGT 136
   | |||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 slvsGlyGlyArgArgGly.....AlaGlyValArgGlyValTyrVal. 59
137 GGTACCGGCTACAAAGCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAG 168
   : |||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 ..PheAlaGlyArgArgLeuSerProArgLys 69
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-557-309B-43
seq_documentation_block:
; Sequence 43, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.4222
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-557-309B-43
```

```
alignment_scores:
Quality: 80.00      Length: 72
Ratio: 1.600        Gaps: 2
Percent Similarity: 69.444      Percent Identity: 26.389
```

alignment_block:

US-09-823-101-4 x US-08-557-309B-43 ..

Align seg 1/1 to: US-08-557-309B-43 from: 1 to: 186

```
289 GCCCAAGACACGTATCAGACTGACGAGAGAAGAAACGACGACGAGG 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 AlarArgrgluAlaGlugluArgrAlaGlInArgrgluAlaGlugluArgrAl 87
339 AGACAACCAACGACGACGCGCGGATGAGCGGACGCGACGACGATACC 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 agInArgrgluAlaGlugluArgrAlaArgrArgrgluAlaGlulysArgrAla 104
389 CGAAGAAACGGGGAAGACACAGATGAAGTAATCTCAGTTGACGAAGAC 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 rgrArgrgluAlaLysgluArgrAlaTrpLn.....GluAlaGlu 116
439 CAAAGCGAGAGACGATGTTGATCAAGAAACCGACGCGACGACGACG 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 GluArgrAlaGlInArgrgluAlaGlugluArgrAlaArgrArgrgluAlaGlu.. 132
489 TGCAGAAGGTGCGAGG 504
||| |||||
133 ....GluArgrAlaArgr 136
```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-834-306-43

seq_documentation_block:

```
: Sequence 43, Application US/08834306
: Patent No. 6054135
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
: NUMBER OF SEQUENCES: 65
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
```

```
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-43
```

```
alignment_scores:
Quality: 80.00      Length: 72
Ratio: 1.600        Gaps: 2
Percent Similarity: 69.444      Percent Identity: 26.389
```

alignment_block:

US-09-823-101-4 x US-08-834-306-43 ..

Align seg 1/1 to: US-08-834-306-43 from: 1 to: 186

```
289 GCCCAAGACACGTATCAGACTGACGAGAGAAGAAACGACGACGAGG 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 AlarArgrgluAlaGlugluArgrAlaGlInArgrgluAlaGlugluArgrAl 87
339 AGACAACCAACGACGACGCGCGGATGAGCGGACGCGACGACGATACC 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 agInArgrgluAlaGlugluArgrAlaArgrArgrgluAlaGlulysArgrAla 104
389 CGAAGAAACGGGGAAGACACAGATGAAGTAATCTCAGTTGACGAAGAC 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 rgrArgrgluAlaLysgluArgrAlaTrpLn.....GluAlaGlu 116
439 CAAAGCGAGAGACGATGTTGATCAAGAAACCGACGCGACGACGACG 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 GluArgrAlaGlInArgrgluAlaGlugluArgrAlaArgrArgrgluAlaGlu.. 132
489 TGCAGAAGGTGCGAGG 504
||| |||||
133 ....GluArgrAlaArgr 136
```


THIS PAGE BLANK (USPTO)


```

409 CATATGAAGTAAGTAATCTCAGTTGACGAACAAAAAGCCGAGACCATGT 458
      |||
501 ValGluuysgluhtststsrGTYrAspGluurghrghstsrGTYrValasphe 517
      |||
459 TCATCTCAAGAAACACACGCGCGGACAGAC 486
      :
517 LgiuSerGluAsnrGAsnrAsnrGserGlu 526

```

seq_documentation_block:
hypothetical protein ZC443.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27574
R:Haynes, C.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z20389
A:Accession: T27574
A:Status: preliminary: translated from GB/EMBL/DDJJ
A:Molecule type: DNA
A:Cross-references: 1-389 <MIL>
A:Cross-references: EMBL:Z75553; PTDN:CAA9946.1; GSPDB:GN00023; CESP:ZC443.4
A:Experimental source: Clone ZC443
C:Genetics:
A:Gene: CESP:ZC443.4
A:Map position: 5
A:Introns: 114/3; 151/3

alignment_scores:	Quality:	98.50	Length:	154
	Ratio:	1.187	Gaps:	5
Percent Similarity:	53.896	Percent Identity:	24.026	
alignment_block:				
US-09-823-101-4 x T27574	..			
Align seg 1/1	to: T27574	from: 1	to: 389	
58	AGCAAAAAAAAAAAAAAAAAAAAAAAAAA	CCGGGGGGAACAGGGA	107	
162	ATgAlAlYserAlYserGInLYsAsnValProAlAtYserPheAsps	178		
108	CATAGGAGATGTCCCGGC	TGGAAC	133
178	proAlAlLeuIleProAlLuGInLYsProIleProGInAlArMrLYsV	195		
134	TGTGTATCCGGCTACAAAATCCCAAGAGAAACATTTCTACAGCGCT	183		
195	aLYrTYrAsnAlValLYsSerValGInAlArProThrProGInAlAla	211		
184	CCACTTCCCGGCACACAGCTAGAGAAACACTTACTAGAAACTACAAAC	233		
212	ProValAlaGInAspValValAlaValAsnGInLeu		222
234	ATATATACATAGTACAGAGAAAGCCCATATGAGTACTCGCATAAAGTA	283		
223	SerGInValLYsLYsTYrCYsGInIleLeu	232
284	CAACTCGCAAAAGACAGTATCAGAGACTGCAGAGAGAGAAACGACAGC	333		
233	LYsAlaGInMeTGLInLYsArgGInLYsArgGInLYs	246	
334	AGAGGAGCAACCAACGAGCCAGCCGTGAGCGCGATGAGCGAGACGCGAG	383		
247	AspAlAlAsnLYsLeuAsnGInValGInLYsLeuIleSe	261	
384	ATATCCGGAAGAAGCGGAGAAAGACAGATGAAGTATTCAGTTGACG	433		
261	rThrAlLYsAspSerCYsSerGIn	LYsG	272
434	AAAGACCAAAAGCAGAAAGACATGTTGATCAAGAAACCGAGCGAGACA	483		

```

|||||.....:||||||| |||... :.....:|||||
272 luAspGluGlutIuAspAspAspAspGluGlutIu 288
484 GACTGTGCAGAA 495
:: .....
289 GluAsnValGlu 292

seq_name: p1r2:T03045
```

```

seq_documentation_block:
hypothetical protein 009R - Chilo iridescent virus
C:Species: Chilo iridescent virus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 04-Mar-2000
C:Accession: F03045
R:Bahr, U.; Tidona, C.A.; Darai, G.
Virus Genes 15, 235-245, 1997
A:Title: The DNA sequence of Chilo iridescent virus between the genome coordinates 0
A:Reference number: Z14834; MUID:98141693
A:Accession: F03045
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-456 <BAH>
A:Cross-references: EMBL:AF003534; NID:g2738385; PIDN:AAB94419.1; PID:g2738392
C:Superfamily: Chilo iridescent virus hypothetical protein 009R

```

```

alignment_block:
US-09-823-101-4 x T03045  ..

Align seg 1/1  to: T03045  from: 1  to: 456

51  TATCAAGAGGAAAAAAAAAAAAAAAAAAAAAAAAACCGGGGGGAAC 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
30  TyrLeuGluLeuLeuGluAsnLysAsnLysValLysLeu..... 42

101 CAGGGACATATGCGAGTGTCCCGGGCGGAAACTGTG..... GTACCCG 144
    43  .....AsnCyS11eGlyLysGluPheValProthThrProP 55
    145 GCTCAAAACTCCAAAAGAGGAACATTTGTA.....ACG 179
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
55  rOlenseThrGlyArgGluThrL1eValSerLysAprOHisVal 71

180 CGCTCCACTACCGCGCACACGCTAGAGAAACATCTACTAGAACTCA 229
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
72  GlnsSerL1eSerAsnThrProL1eArgThrL1eLysAspThrP 88

230 CAACATATATACATATAGTACAGAGAAAGCCACATAGAGTCATCCGATAA 279
    88  oArg.....TyrGluGluThrProL1eLysArgThrL1eThy 101
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
280  AGTACAACTGC.....GCA 293
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101  alThrsValLysThrValLysSerSerSerL1eSerGlyMetAsnGly 117

294 AAGACAGCTATCAGAGACTGACGAGAGAGAAACGACGACAGAGAGACA 343
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
118  ArgAsnArgLeuTyrAspAspAspLeuPheAspAspArgTyrLys 134

344 AACCAAGACGCCACCGT.....GAGCGCATAGGACGACAGCCG 381
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134  erProThrThrArgLysPheGlnGlyLysArgAspGluAspL1eArg 150

382 AGATPACCGGAGAGAAAG.....GGGAAAGACAGATGAAATATC... 423
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151  LeuL1eProLysSerSerAsnL1eGlySerSerLysTyrLysProValL 167

424  ....TCACTTACGACAGACCAAAAGCGAGAGACGATGTTGATCAAAGA 469

```


seq_documentation_block:
hypothetical protein Y38E10A.f - *Caenorhabditis elegans*

431 ACGAAGACCAAAAGCGAGAAGACGAT..... 45

seq_documentation_block:
hypothetical protein Y38E10A.f - *Caenorhabditis elegans*


```

66 AAAAAAAAAAAAAAAAAA.....AAACCCGGGGGA 97
   |||
96 sthrvalthrproalalysalavalthrthrprogllysllysglyalatr 113
   |||
98 AACCGGGGAC...ATACGAGTCTCCGGCGTGGAA...CTGGGTAC 141
   |||
113 hrprogllyslalaleuvalalatrhrprogllysllysglyalalale 129
   |||
142 CCGGTACAAACCTCCAAAGAGAAACATGTCAACGGGCTCCACTACC 191
   |||
130 proalalysgllyalalysasmelyllysasnalalyslysgluaspsers 146
   |||
192 GCGCACACAGCTAGAGAAACACATCTACTAGAACTCAACATATATAC 241
   |||
146 poluglugluaspaspaspsercglugluaspugluaspaspaspasp 163
   |||
242 ATAGTACAGAGAAACGCCACATAGACTACTGCAATAAAGTCAACTGCG 291
   |||
163 luaspgluaspgluaspgluileggluoproalaalameetylalalala 179
   |||
292 CAAGACACGTTATCAGACGTGACGAGAGAGAAAGACGACAGAGAGA 341
   |||
180 AlaAlaAlaProAlaSerGlu...AspGluaspaspGluaspaspGluasp 195
   |||
342 CAAGCAACGACGACGCGTACGCGATGAGCGACAGCGCAATACCCGA 391
   |||
195 poluaspaspaspaspaspaspglugluaspaspsercglugluAlameG 212
   |||
392 AAGAAACG.....GGGAAAGACAGATGAAGTAACTCTCACTT... 429
   |||
212 luThrThrProAlaLysgllysllyslalalalysvalalalProValLys 228
   |||
430 .....GACGAAGACCAAAAGCAGA 449
   |||
229 AlaLysasnValAlaGluaspGluaspGluaspGluaspGluaspGlu 245
   |||
450 AGACGATGTTGATCAAAAGAAACACGCGCAGACAGAC 486
   |||
245 uaspaspaspaspaspaspGluaspaspGluaspaspasp 257

seq_name: p1r2:JE0378

seq_documentation_block:
DNA (cytosine-5)-methyltransferase (EC 2.1.1.37) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0378
R:Kimura, H.; Takeda, T.; Tanaka, S.; Ogawa, T.; Shiota, K.
Biochem. Biophys. Res. Commun. 253, 495-501, 1998
A:Title: Expression of rat DNA (cytosine-5) methyltransferase (DNA MTase) in rodent trop
A:Reference number: JE0378; MUID:99097263
A:Accession: JE0378
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1622 <KIM>
A:Cross-references: DDBJ:AB012214; NID:q4160669; PIDN:BA37118.1; PID:q4160670
C:Keywords: methyltransferase; S-adenosylmethionine

```

```

alignment_scores:
  Quality: 96.50      Length: 149
  Ratio: 1.109      Gaps: 10
  Percent Similarity: 58.389      Percent Identity: 29.530

```

alignment_block:

US-09-823-101-4 x JE0378 ..

Align seg 1/1 to: JE0378 from: 1 to: 1622

```

57 GAGGAAAAAAAAAAAAAAAAAAAAAAAAAACCAGGGG...GAACACAG 103
   |||
191 GluGluArgAspGluAspLysLysArgValAlaGlyThrGluSerAr 207

```

```

104 G.....GGACATAGC...GAGTGTCCCGGCTGGAAACTGTGG 138
   |||
207 galaserarglialglyluservalglulyspro.....GluArgValA 222
   |||
139 TACCCGCTACAAACCTCCAAAGAGAAACATTTTCAACGCGCTCCACT 188
   |||
222 rgProglYThrGlnLeucysgInglugInlu...GlyGluGlnGluasp 237
   |||
189 ACCGGGCACACAGCTAGAGAAACACATCTACTAGAAACCTCAACATATA 238
   |||
238 AsparGArgProArgArgGlnThrArgGlnLeuAlaSerArgArgLys.. 253
   |||
239 TACATAGTACAGAGAAACGCGACATAGAGCTACTGCAATAAAGTCAACT 288
   |||
254 .....SerArgGluaspProAspArgGlnLarPro.GlyThr... 266
   |||
289 GCGCAAGACACGTTATCAGAGACTGACGAGAGAGAAACGACGACAGAG 338
   |||
267 .....HisLeuAspValAspaspaspaspGluLysAspLysArgSe 280
   |||
339 AGACAACCAACGACGCGCGTACGCGATGAGCGCAGCGCAGATAC 388
   |||
280 rSerArgProArgSerGln...ProArgAspLeuAlaThrLysArgArgp 296
   |||
389 CGAAAGAAACGCGGAAAGACAGATGAAGTAACTCACTTGCAGCAAGAC 438
   |||
296 roLysGlu.....GluValGluGlnIleThrProGluProPro 308
   |||
439 CAAGACGAGAGACGATGTTGATCAAGAAACACAGCGGAGACA 483
   |||
309 GluGlyLysAspGluaspGluArgGluGluLysArgArgLysThr 323

seq_name: T19201

seq_documentation_block:
hypothetical protein C11G6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T19201
R:Wilkinson, J.
Submitted to the EMBL Data Library, March 1996
A:Reference number: Z19089
A:Accession: T19201
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-385 <WIL>
A:Cross-references: EMBL:Z70204; PIDN:CAA94113.1; GSPR3:GN00028; CESP:C11G6.3
A:Experimental source: clone C11G6
C:Genetics:
A:Gene: CESP:C11G6.3
A:Map position: X
A:introns: 12/1; 106/3; 171/2; 335/2

```

```

alignment_scores:
  Quality: 96.00      Length: 137
  Ratio: 1.247      Gaps: 4
  Percent Similarity: 56.204      Percent Identity: 21.898

```

alignment_block:

US-09-823-101-4 x T19201 ..

Align seg 1/1 to: T19201 from: 1 to: 385

```

76 AAAAAAAAAAAACCGGGGGGAAACAGGGGACATAGCGAGTGTCCGGG 125
   |||
87 LysLysGluGluProSerSerMetThrProGlu...SerSerArgProGl 102
   |||
126 CTGGAACACTGTGTACCGGCGTACAAACCTCCAAAGAGGAACATGTGC 175
   |||
102 ySerSerLeuGluThrProSerSerSerSerSerLysHisHisHisH 119

```



```
119 InLeuArgGluGluGluGluTyrLysArgGluLeuLeuAlaGluArgGln 135
277 .....TAAAGTACAACTGCCCAAGACACAGTATCAGAGACTGA 314
136 LysArgIleGluGluGluGluLysGluGluArgArgArgLeuGluGluGln 152
315 CGAGAGAAGAAACGACGACAGAGAGACCAACCAAGCCAGCGGTGAGC 364
152 narArgGluArgGluAlaArgArgGluGluArgGluGluArgArgA 169
365 GCGATGAGGCACAGCGCAGATACCCGAAGAAGACGGGAAGAAGACAGATG 414
169 rgGluGluGluGluLysArgArgLeuGluLeuGluArgArgArgLys 185
415 AAAGTAATCTCAGTTGACCAAGACCAAAAGCGAAGACGATGTGATCA 464
186 Glu.....GluGluLysArgArgAlaGluGluLysAr 198
465 AAGAAACGACGCCGAGACAGAC 486
198 garGValGluArgGluGlu 205
```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-4 to: SwissProt_39:* out-format : pfs

Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ n2p.model -DEV-xlp
-O-qcgn2.1/USPTO.spool/US09823101/runat_17012002.155138.22061/app-query.fasta_1.2239
-DB-SwissProt_39 -OPMT-fastan -SUFFIX-n2p.rsp -GAPEXT-12.000
-GAPEXT-4.000 -MINMATCH=0.100 -LOOPEXT-0.000 -LOOPEXT-0.000
-QCAPOP=4.500 -QCAPEXT=0.050 -XGAPOP=10.000 -XCAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-biosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USPTO-0509823101.ccgnt_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-MAIL -THREADS=1

Search information block:

Query: US-09-823-101-4
Query length: 509
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 66.470000

score_list:

Sequence	Strid	Orig	Zscore	EScore	Len	Documentation
SwissProt_39:MTDM_MOUSE	110	00	166.14	0.0086	1620	P13864 mus musculus (mouse)
SwissProt_39:YKR2_YEAST	97	00	156.86	0.1390	427	P36049 saccharomyces cerevisiae
SwissProt_39:NUCL_HUMAN	96	50	150.03	0.1562	706	P19393 homo sapiens (human)
SwissProt_39:XPR6_YARLI	95	00	144.94	0.2170	976	P42781 yarrowia lipolytica
SwissProt_39:TFHY_SHEEP	94	50	140.44	0.2437	1549	P22793 ovina aries (sheep)
SwissProt_39:TRD3_HUMAN	94	00	142.43	0.2697	1083	P00068 homo sapiens (human)
SwissProt_39:SNL1_YEAST	94	00	136.70	0.2731	2231	P00416 saccharomyces cerevisiae
SwissProt_39:SPR8_HUMAN	92	00	145.46	0.4094	484	P00519 homo sapiens (human)
SwissProt_39:NUCL_MESAU	91	50	141.55	0.4590	713	P08199 mesocricetus auratus
SwissProt_39:GAG_MSVMO	91	00	142.94	0.5088	538	P03334 moloney murine sarcoma
SwissProt_39:INCE_CHICK	91	00	139.07	0.5131	877	P53352 gallus gallus (chicken)
SwissProt_39:GAG_MLVMO	90	50	142.07	0.5668	540	P21435 homio murine leukemia
SwissProt_39:YUJ2_CAEEL	90	00	145.79	0.6250	304	P33836 caenorhabditis elegans
SwissProt_39:ATRX_HUMAN	89	00	127.80	0.8036	2375	P46100 homo sapiens (human)
SwissProt_39:MT2_DROHY	88	50	131.20	0.8868	1391	P08696 drosophila hydei (fruit fly)
SwissProt_39:NM5M_TRYBB	88	00	137.17	0.9731	590	P04540 trypanosoma brucei (tsetse fly)
SwissProt_39:IF3A_MAZE	88	00	133.29	0.9814	962	P04540 trypanosoma brucei (tsetse fly)
SwissProt_39:SPR8_HUMAN	87	50	133.54	1.09	951	P01872 zea mays (maize)
SwissProt_39:CCAB_DDISM	87	50	125.44	1.11	2326	P12872 homo sapiens (human)
SwissProt_39:NMN4_YEAST	87	00	130.00	1.122	1178	P36698 saccharomyces cerevisiae
SwissProt_39:IF2P_HUMAN	87	00	129.72	1.122	1120	P06841 homo sapiens (human)
SwissProt_39:AFSA_EHENT	87	00	127.20	1.123	1676	P00083 emericella nidulans
SwissProt_39:TF2A_HUMAN	86	50	135.69	1.134	517	P35269 homo sapiens (human)
SwissProt_39:IF3A_MAZE	86	50	135.38	1.134	538	P03332 moloney murine leukemia
SwissProt_39:GAG_MLVMO	85	50	133.76	1.167	605	P13916 glycine max (soybean)
SwissProt_39:GAGA_SOYBN	85	50	126.07	1.169	1407	P37709 oryctolagus cuniculus
SwissProt_39:PHRY_RABIT	85	50	118.05	1.172	3866	P55200 mus musculus (mouse)
SwissProt_39:HRX_MOUSE	85	00	126.54	1.188	1192	P046072 drosophila melanogaster
SwissProt_39:K2_DROME	85	00	126.07	1.189	1265	P06431 mus musculus (mouse)
SwissProt_39:MYO6_MOUSE	85	00	124.66	1.189	1523	P18583 homo sapiens (human)
SwissProt_39:SON_HUMAN	84	50	132.04	2.07	536	P29168 murine leukemia virus
SwissProt_39:GAG_MLVMO	84	50	132.04	2.07	539	P26807 friend murine leukemia
SwissProt_39:GAG_MLVMO	84	50	131.38	2.07	583	P26044 sus scrofa (pig)
SwissProt_39:RDI_PIG	84	50	130.21	2.07	675	P03514 ratius norvegicus (rat)
SwissProt_39:GLYB_PLAFG	84	50	129.13	2.08	774	P02895 plasmodium falciparum
SwissProt_39:TOP1_XENLA	84	50	128.55	2.08	829	P41512 xenopus laevis (african clawed toad)
SwissProt_39:YUJ1_YEAST	84	00	124.55	2.34	1240	P53335 saccharomyces cerevisiae
SwissProt_39:MYO6_HUMAN	84	00	124.41	2.34	1262	P09054 homo sapiens (human)
SwissProt_39:SCG7_YEAST	83	50	119.88	2.62	2009	P11075 saccharomyces cerevisiae
SwissProt_39:TOD2_CAEEL	83	00	136.13	2.81	233	P34594 caenorhabditis elegans

SwissProt_39:IF3A_TORAC + 83.00 124.92 2.88 958 P040554 nicotiana tabacum
SwissProt_39:GLT_DROME + 83.00 124.39 2.89 1023 P33438 drosophila melanog
SwissProt_39:HRP1_SCHPO + 83.00 122.06 2.90 1373 P09525 schizosaccharomyc
SwissProt_39:CG7_HUMAN + 82.50 132.04 3.16 351 P09388 homo sapiens (huma
SwissProt_39:CBF5_KULIA + 82.50 129.66 3.17 474 P013473 kluyveromyces lact

seq_name: SwissProt_39:MTDM_MOUSE

seq_documentation_block:
ID MTDM_MOUSE STANDARD; PRT; 1620 AA.
AC P13864;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA (CYTOSINE-5)-METHYLTRANSFERASE MMU1 (EC 2.1.1.37) (DNA
DE METHYLTRANSFERASE MMU1) (DNA MTASE MMU1) (MCM1) (M.MMU1).
GN DNMT1 OR DNMT OR UDM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69094873; Pubmed=3210246;
RA Bestor T.H., Laudano A., Mattaliano R., Ingram V.;
RT "Cloning and sequencing of a cDNA encoding DNA methyltransferase of
RT mouse cells. The carboxyl-terminal domain of the mammalian enzymes is
RT related to bacterial restriction methyltransferases.";
RL J. Mol. Biol. 203:971-983(1988).
RN [2]
RP REVISIONS TO N-TERMINUS.
RC TISSUE=Embryo;
RX MEDLINE=97094871; Pubmed=8940105;
RA Yoder J.A., Yen R.-W.C., Vertino P.M., Bestor T.H., Bayliss S.B.;
RT "New 5' regions of the murine and human genes for DNA (cytosine-5)-
RT methyltransferase.";
RL J. Biol. Chem. 271:31092-31097(1996).
CC -1- FUNCTION: METHYLATES CG RESIDUES.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA -> S-ADENOSYL-L-
CC HOMOCYSTEINE + DNA CONTAINING 5-METHYLCYTOSINE.
CC -1- SIMILARITY: HIGH, TO OTHER EUKARYOTIC DNA METASE.
CC -1- SIMILARITY: SOME TO BACTERIAL RESTRICTION SYSTEMS.
CC METHYLTRANSFERASES.
CC -1- SIMILARITY: CONTAINS 2 BAH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; X14805; CA32910.1; -;
DR PIR; S01845; S01845.
DR REBASE; 2844; M.Mmu1.
DR MGD; MGI:94912; Dnmt1.
DR InterPro; IPR001025; BAH.
DR InterPro; IPR001525; C5_DNA_meth.
DR InterPro; IPR002857; ZNF-CXXC.
DR Pfam; PF01426; BAH; 2.
DR Pfam; PF00145; DNA_methylase; 3.
DR Pfam; PF02008; ZF-CXXC; 1.
DR PRINTS; PR00105; C5METRFRASE.
DR SMART; SM00439; BAH; 2.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
DR TRANSFERASE; Methyltransferase; DNA-binding.
KW DOMAIN 651 693 CYS/ARG/LYS-RICH.
FT AC SITE 1229 1229 BY SIMILARITY.
SQ SEQUENCE 1620 AA; 183286 MW; F73710AD043E709 CRC64;

alignment_scores: Quality: 110.00 Length: 148
 Ratio: 1.325 Gaps: 7
 Percent Similarity: 56.081 Percent Identity: 28.378

alignment_block:

US-09-823-101-4 x MTDN_MOUSE ..

Align seg 1/1 to: MTDN_MOUSE from: 1 to: 1620

```

54 CAAGAGGAGAAAAA.....AAAAAAAAA.....CCCGGGGGAACAG 103
   :::::..... |||||.....: ||:::
192 GluGluArgSpGlnAspLysArgValValAspThrGlnSerG1 208
   :::::..... |||||.....: ||:::
104 GGGACATAGCGAGTGTCCCGGCTGGAACCTGTGTACCGGCTAC.... 149
   :::::..... |||||.....: ||:::
208 YAlAlAlAlAlAlAlValGluLysLeuGluValThrAlaGlyThrGln 225
   :::::..... |||||.....: ||:::
150 .....AAACTCCAAAGAGGAACATTGTCAACGGCCTCCACTAC 191
   :::::..... |||||.....: ||:::
225 euGlyProGluGluProCysGluGluGlnLysAspAsnArgSerLeuArg 241
   :::::..... |||||.....: ||:::
192 GCGACACAGCTAGAGAAACATCTACTAGAAACTCACAACATATATAC 241
   |||||..... |||||.....: ||:::
242 ArgHisThr...ArgGluLeuSerLeuArgArgLysSer..... 253
   :::::..... |||||.....: ||:::
242 ATAGTACAGAGAAAGCCACATAGCTACTGCATTAAGTAACTGCG 291
   :::::..... |||||.....: ||:::
254 .....LysGluAspProAspArgGluAlaArgProGluThr..... 265
   :::::..... |||||.....: ||:::
292 CAAGACAGCTATCAGAGACTGACAGAGAGAAGAACGACAGAGAGA 341
   |||||..... |||||.....: ||:::
266 .....HisLeuAspGluAspGluAspGlyLysLysAspLysArgSer 280
   :::::..... |||||.....: ||:::
342 CAAGCAACAGACGCGAGCGTGAAGCGGAGAGCCAGATACCCGA 391
   :::::..... |||||.....: ||:::
280 ArgProArgSerGln...ProArgAspProAlaAlaLysArgArgPro 296
   :::::..... |||||.....: ||:::
392 AAGAAACGGGGGAAACAGACAGATGAAGTAATCTCAGTTGACGAAGCA 441
   |||||..... |||||.....: ||:::
296 YsgLysValArgAlaArgAlaGlySer.....SerArgAspSerGluAsp 310
   :::::..... |||||.....: ||:::
442 AAGCGAGAGACGATGTTGATCAAGAAACGAGCGGAGACA 483
   :::::..... |||||.....: ||:::
311 ArgAspGluAspGluArgGluGluLysArgArgLysThrThr 324
   :::::..... |||||.....: ||:::

```

seq_name: SwissProt_39:YKR2_YEAST

seq_documentation_block:

```

ID YKR2_YEAST       STANDARD:       PRT:   427 AA.
AC P36049;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHELTICL 49.7 KDA PROTEIN IN GIN2-STE3 INTERGENIC REGION.
GN YKL172W OR YKL636.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378719; PubMed=8091858;
RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "Sequencing and analysis of a 20.5 kb DNA segment located on the left
   arm of yeast chromosome XI.";
RL Yeast 10:S25-S33(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

DR EMBL: Z26878; CA81515.1; -.
DR EMBL: Z28172; CA82014.1; -.
DR PIR: S38002; S38002.
DR PIR: S38409; S38409.
DR PIR: S44589; S44589.
DR SGD: S0001655; YKL172W.
KW Hypothetical protein.
SQ
SEQUENCE 427 AA; 49734 MW; 4A11F6CDE779DB5A CRC64;

```

alignment_scores: Quality: 97.00 Length: 161
 Ratio: 1.102 Gaps: 9
 Percent Similarity: 54.658 Percent Identity: 24.845

alignment_block:

US-09-823-101-4 x YKR2_YEAST ..

Align seg 1/1 to: YKR2_YEAST from: 1 to: 427

```

37 AAAAAAGTATTCTATCAAGAGGAAAAA.....AAAAAAA 83
   ||| :::::..... |||||.....: |||||
22 LysLeuGluAsnAspLeuLysLysLysSerGlnGluLeuLysLysG1 38
   :::::..... |||||.....: |||||
84 AAAACCCGGGGGAAACCAAGGACATAGCGAGTGTCCCGGCTGGAAC 133
   :::::..... |||||.....: |||||
38 uGluPro.....ThrIleValThrAlaSerAsnLeuLys. 49
   :::::..... |||||.....: |||||
134 TGtGTATCCCGGCTACAAACTCCAAAGAGGAACATTGTCAACGGCT 183
   |||||..... |||||.....: |||||
50 .....LysLeuGluLysGlu..... 55
184 CCACTACCGCGCACACAGCTAGAGAAACATCTACTAGAACTACAAAC 233
   :::::..... |||||.....: |||||
56 .....LysLysAlaAspValLysLysGluValAlaAlaAspThrGlu 70
   :::::..... |||||.....: |||||
234 ATAT.....ATACATAGTACAGAGAAACCCACATA.....G 265
   ||| :::::..... |||||.....: |||||
70 uTyrGlnSerGlnAlaLeuSerLysLysGluLysArgLysLeuLysG 87
266 AGCTATCTC.....GCATTAAGTCACTGCGCAAGACACGTA 303
   ||| :::::..... |||||.....: |||||
87 LysLeuLysLysMetGlnGluGlnAspAlaThrGluAlaGlnLysHisMet 103
   :::::..... |||||.....: |||||
110 rGlyAspAspArgGluGluGluGluGluGluGluGluGluGlyA 127
404 AAAGACAGTAAGTAATCTCAGTTGACGAGAACCAAGCAAGAAC 453
127 rGluAspLeuGluLysLeuAla...LysSerAspSerGluSerGluAsp 142
454 GATGTTGATCAAGAAACGAGCGGAGACAGAC 486
   ||| :::::..... |||||.....: |||||
143 AspSerGluSerLysLysAspSerGluGluAsp 153
seq_name: SwissProt_39:NUCL_HUMAN
seq_documentation_block:
ID NUCL_HUMAN       STANDARD:       PRT:   706 AA.
AC P19338;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEOLIN (PROTEIN C23).

```

GN NCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9290043; PubMed=2737305;
RA Strivastava M., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Cloning and sequencing of the human nucleolin cDNA."
RL FRS Lett. 250:99-105(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9036866; PubMed=2394707;
RA Strivastava M., McBride O.W., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Genomic organization and chromosomal localization of the human
nucleolin gene."
RL J. Biol. Chem. 265:14922-14931(1990).
CC -1- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING
EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR
CHROMATIN AND PERIBIOSOMAL PARTICLES. IT INDUCES CHROMATIN
CONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A
ROLE IN PRE-RRNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -1- SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M60858; AAA5954.1; -
DR PIR: A35804; A35804.
DR PIR: S04631; S04631.
DR HSP: P09651; IUP1.
DR Aarhus/Ghent-2DPAGE; 1210; NEPHGE.
DR MIM: 164035; -
DR InterPro: IPR00504; RRM.
DR Pfam: PF00076; TTM; 4.
DR SMART: SM00360; RRM; 4.
DR PROSITE: PS00102; RRM; 4.
DR PROSITE: PS00030; RRM_RNP_1; 3.
KM Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;
RNA-binding.
FT INIT_MET .0
FT DOMAIN 142 170 ASP/GLU-RICH (ACIDIC).
FT 184 208 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 233 270 ASP/GLU-RICH (ACIDIC).
FT 306 382 RNA-BINDING (RRM) 1.
FT DOMAIN 392 465 RNA-BINDING (RRM) 2.
FT 485 559 RNA-BINDING (RRM) 3.
FT DOMAIN 571 643 RNA-BINDING (RRM) 4.
FT 645 694 ARG/GLY/PHE-RICH.
FT 57 134 8 x 8 AA TANDEM REPEATS OF X-T-P-X-K-K-X-X.
FT DOMAIN
FT REPEAT 57 64 1.
FT REPEAT 74 81 2.
FT REPEAT 82 89 3.
FT REPEAT 90 97 4.
FT REPEAT 97 103 5.
FT REPEAT 104 111 6.
FT REPEAT 119 126 7.
FT REPEAT 127 134 8.
FT MOD_RES 144 144 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 152 152 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 183 183 PHOSPHORYLATION (BY SIMILARITY).
SO SEQUENCE 706 AA; 76213 MW; 85A2F2CA22EA03DB CRC64;

alignment_scores:

Quality: 96.50 Length: 179
Ratio: 1.027 Gaps: 6
Percent Similarity: 52.514 Percent Identity: 24.022

alignment_block:
US-09-823-101-4 x NUCL_HUMAN ..

Align seg 1/1 to: NUCL_HUMAN from: 1. to: 706

```

16 AGGATGCTGATATATCATCATGAGGAAAAACGTAATCTATCAAGAGGAAAA 65
   :::::::::::::: ||| ||| |||
79 LysAlaAlaValThrProGlyLysAlaAlaAlaThrProAlaLysLys 95
   AAAAAAAAAAAAAAAAAA.....AAACCGGGGGA 97
   | |||
95 sThrValThrProAlaLysAlaValThrThrProGlyLysLysGlyAlaT 112
   |||||
98 AACCGAGGAGC...ATACGAGCTCCCGGGCTGAAA...CTGTGTC 141
   |||||
112 hrProGlyLysAlaLeuValAlaThrProGlyLysLysGlyAlaAla 128
   |||||
142 CCGGCTACAAACCTCCAAAGAGAAACATGTGTCAAGCGGCTCCACTACC 191
   |||||
129 ProAlaLysGlyAlaLysAsnGlyLysAsnAlaLysLysGlyLysSerAs 145
   ::::
192 GCGCACACGCTAGAGAAACACATCTACTAGAACTCAACATATATATAC 241
   ::::
145 poluIuIuLysAspAspAspSerGluGluAspGluLysAspGluAspG 162
242 ATAGTACAGAGAAACGCCATAGACTACTCGCATTAAGTACAACTCGC 291
   ::::
162 LuAspGluAspLysAspLysGluIleGluProAlaAlaMetLysAlaAla 178
222 CAAGACACGCTATCAGACGTCAGAGAGAAAGAACGACGACGAGGAGA 341
   ::|||
179 AlaAlaProAlaSerGlu...AspGluAspAspGluAspAspGluAspAs 194
342 CAAGCAACGAGCCAGCGCTAGCGGATGAGCAGAGCGCATATATATAC 391
   ::::
194 poluAspAspAspAspAspGluGluAspAspSerGluGluAlaLeuMet 211
392 AAGAAACG.....GGGAAAGACAGATGAAGTATCTCACTT... 429
   ::|||
211 LuThrThrProAlaLysGlyLysLysAlaAlaLysValProValProLys 227
430 .....GACGAAAGCCAAAGCAGGA 449
228 AlaLysAsnValAlaGluAspGluAspGluGluGluAspGluAspGlu 244
450 AGACGATGTTGATCAAGAAACACGCGGAGAGACAGAC 486
   |||||
244 uAspAspAspAspAspGluAspGluAspGluAspAsp 256

```

seq_name: SwissProt_39:XPR6_YARLI

seq_documentation_block:
ID XPR6_YARLI STANDARD; PRT; 976 AA.
AC P42781;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIBASIC PROCESSING ENDOPROTEASE PRECURSOR (EC 3.4.21.-).
GN XPR6.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CX161-1B;
RX MEDLINE=94262316; PubMed=8203153;
RA Enderlin C.S., Ogrzydzak D.M.;
RT "Cloning, nucleotide sequence and functions of XPR6, which codes for

[illegible]

```

FT VARIANT 1145 1197 MISSING (IN SHORT FORM).
FT VARIANT 1251 1273 MISSING (IN SHORT FORM).
FT CONFLICT 1399 1399 E->C (IN REF. 2).
SQ SEQUENCE 1549 AA: 201173 MW: EYZB9FFJ326E54E CRC64;

alignment_scores:
    Quality: 94.50 Length: 126
    Ratio: 1.312 Gaps: 4
Percent Similarity: 57.143 Percent Identity: 24.603

alignment_block:
US-09-823-101-4 x TRHY_SHEEP ..

Align seg 1/1 to: TRHY_SHEEP from: 1 to: 1549

127 TGGAAACTGTGTACCGGCTCAAAACTCCAAAGAGGAAACATTGTCA 176
128 ..:::|||||:::
365 TRPATG...TTPGlnLeuGlnGlnIleSerGlnAArgArgTyrThrLe 380
177 ACGGCT...CCACTACGGCGACACAGCTAGAGAACAACATCTACTGA 223
178 ||| ||| ||| |||:::
380 UTYrAlaIysProAlaGlnArgGlnGlnValArgGlnGlnGlnLeuA 397
224 AACTGCACAACATATATATACATGTACAGAGAAAGCCACATAGAGCTACTC 273
225 ::||| ::|||:::
397 TgLeuIySgIuGlnIuIySLeuGlnAArgIuIySArgArgGlnIu..... 411
274 GCATTAAGTACAACTGGCCGAACACGATATCAGACAGCTGACAGAGAAG 323
412 .....ArgIuIySgIuIySArgIuIySArgIuIySArgIuIySArgIu 423
324 AAACGACGACAGAGAGAGACAAACACACAGCCAGCGCTGAGCGATGAGG 373
423 |:::|||||::: |||::: |||::: |||::: |||
423 gIuIuGlnIuIySArgIuIySLeuIuIySArgIuIySLeuIuIySArgIuIyS 440
374 CAGACGCGCATACCCGAAAGAAACGGGGAAGAAAGACAGTGAATATC 423
440 |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
440 TgIuIySArgIuIySArgIuIySArgIuIySArgIuIySArgIuIySArgIuIyS 456
424 TCAGTT.....GACCAAGACCAAAAGCGAGAGACGATGTGATCAAG 467
425 ::|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
457 GluLeuTTPgIuIuGlnIuIySLeuIuIySArgIuIySArgIuIySArgIuIyS 473
468 AAACGACGCGGACGACAGACACTGTGCAGAA 495
473 gGlnIuIySArgIuIySArgIuIySArgIuIySArgIuIySArgIuIySArgIu 482

seq_name: Swissprot_39:T2D3_HUMAN

seq_documentation_block:
ID T2D3_HUMAN STANDARD; PRT; 1083 AA.
AC 000268; Q99721;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION INITIATION FACTOR TFIID 135 KDA SUBUNIT (TAFTII-135)
DE (TAFTII135) (TAFTII-130) (TAFTII130).
GN TAF2C1 OR TAF2C OR TAFTII135 OR TAFTII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RX MEDLINE=97336072; Pubmed=9192867;
RA Meguid G., May M., Carre L., Chambon P., Davidson I.:
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells.";
RL Genes Dev. 11:1381-1395(1997).
RP [2]
SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.

```

```

RX MEDLINE-97098442; PubMed-8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIIID
RT complex: htraf130 and htraf100".
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y11354; CAA72189.1; -
DR EMBL: U75308; AAC50901.1; -
DR HSSP: P19656; IAFH.
DR MIM: 601796; -
DR InterPro: IPR003894; TAF_hom.
DR SMART: SM00549; TAFH; 1.
KM Transcription regulation; Nuclear protein.
FT DOMAIN 39 42 POLY-HIS.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 142 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT DOMAIN 105 117 PGPSPRRLVPA -> GRGLDQGRGRES
FT CONFLICT 105 117 (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 2).
FT CONFLICT 293 293 P -> L (IN REF. 2).
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

alignment_scores:
Quality: 94.00 Length: 228
Ratio: 0.969 Gaps: 13
Percent Similarity: 42.544 Percent Identity: 23.246

alignment_block:
US-09-823-101-4 x T2D3_HUMAN ..

Align seg 1/1 to: T2D3_HUMAN from: 1 to: 1083
67 AAAAAAAAAAAAAAAAAACCGGGGGAACCA..... 102
:::|||||:|||||:|||||:|||||:|||||:
814 Glnltsnlnlsleuylsgluprpglyglyserhearfgaspas 830
103 .....GGGGACATAGCGAGTGTCCGGCGTGAACCTGTGTAACCGG 145
:::|||||:|||||:|||||:|||||:|||||:
830 pasplleasnspvalalasermetaglvalasnleuserglus 847
146 CTACAAACAC.....CAAAAGAGGAAA 168
847 erlaaaglleuvalaThAsnsergluleuvalglyThrleuTharg 863
169 CATGTGCA.....CGCGTCACACTACCGCGCACACA 200
|||||:|||||:|||||:|||||:|||||:
864 serCysylsaspgluThrpheluLeuGlAlaAlaProleuGlnarGatg 880
201 GCTAGAG.....AAACAC.....ATCTACTAGAACTCA 229
|||||:|||||:|||||:|||||:

```

```

880 eleuGluleGlyLysLysHisGlyIleThrgluLeuHisProAspVal 897
230 CAACATATATA...CATAGTACAGAAACGCCACATAGAGTACATCGCA 276
:::|||||:|||||:|||||:|||||:|||||:
897 alserTYValSerHisAlaThrcInlArgluGlnAsnleuVal... 912
277 TAAAGTCAACCTGGCGCAACACGTATCAGACACTGACGAGAAAGAA 326
913 .....GlnylIleSerGluThrAlaGlnInlLysas 923
327 C.....GACGACAGAGGACAAACACACGACGCT. 360
923 nphserTYLysAspAspArgTYrGluGlnAlaSerAspValArgA 940
361 .....GAGCGC 366
940 lagInleuLysPhepneGluGlnleuAspClnIleGluLysGlnArgLys 956
367 GATGAGGACAGAGCGCAGATACCGAAAGAAACGGGAAAGACAGATGAA 416
957 AspGluGlnArg.....GluIleLeuMetArgAlaAla 969
417 AGTATCTCAGTTGACGACAC..... 438
969 sserArgSerArgGlnGlnAspProGluGlnleuArgLysGlnLysA 986
439 .....CAAAAGCGAAGACGAT... 456
986 lalysGluMetGlnGlnGlnleuAlaGlnMetArgGlnArgAspAla 1002
457 .....GTTGATCAAAAGAAACCGCGGACGACAGAGA 485
1003 AsnleuThrAlaLeuAlaIleGlyProArgLysLysArgLysValAs 1019
486 CTGT.....GCACAGGTGGC 501
1019 pcysProGlyProGlySerGlyAlaGlnGlySer 1030

seq_name: SwissProt_39:SENI_YEAST
seq_documentation_block:
ID SENI_YEAST STANDARD; PRT; 2231 AA.
AC Q00416; Q06448;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE TRNA-SPlicing ENDONUCLEASE POSITIVE EFFECTOR.
OS SENI OR YER430W OR Y9576.1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
[1]
RN RP SEQUENCE OF 120-2231 FROM N.A.
RX MEDLINE-92236590; PubMed-1569945;
RA Demarini D.J., Winey M., Ursic D., Webb F., Culbertson M.R.;
RT "SENI, a positive effector of trna-splicing endonuclease in
RL Mol. Cell. Biol. 12:2154-2164(1992).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Meneses S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Tatch A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR ENDONUCLEOTIC CLEAVAGE OF INTRONS FROM
CC ALL FAMILIES OF PRECURSOR TRNAS. MAY BE ONE OF SEVERAL COMPONENTS
CC OF A NUCLEAR-LOCALIZED SPLICING COMPLEX. SENI IS ESSENTIAL FOR

```


SQ SEQUENCE 713 AA: 76997 MW: 79DDCF724CED7B4 CRC64;

Alignment_scores:

Quality: 91.50 Length: 170
Ratio: 0.943 Gaps: 7
Percent Similarity: 57.059 Percent Identity: 24.118

Alignment_block:
US-09-823-101-4 x NUCL_MESAU ..

Align seg 1/1 to: NUCL_MESAU from: 1 to: 713

```

38 AAAAAAGTAAATCTATCAAGAGAAAAA.....AA 69
|||||:|||||:|||||:|||||
60 LysLysValValValSerGlnThrLysLysValAlaValProThrProAl 76
70 AAAAAAAAAAAAAACCGGGGGGAAACCGGACATAGCGAGTGT 119
|||||:|||||:|||||:|||||
76 aLysLysAlaAlaValThrProGlyLysLysAlaAlaAlaThrProAl 93
120 CCGGGGGGAAAGTGTGTACTGCGGCTACAAACTCCA...AAGAGGA 166
|||||:|||||:|||||:|||||
93 yLysAla.ValThrProAlaLysAlaValAlaThrProGlyLysLysG 109
167 AACATTGTCACGCGCTCCACTACCGCGCACAGCTAGAGAAACACATC 216
|||:|||||:|||||:|||||
109 yAlaThrGlnAlaLysAlaLeuValAlaThrProGlyLysLysGlyAla 126
217 TACTAGAAACTCACACATATATATAGTACAGAGAAACCGCACATAGA 266
|||:|||||:|||||:|||||
126 A.....ThrProAlaLys 130
267 GCTACTCGCATAAAGTACACTGGCGCAAGACAGTATCAGAGCTGACG 316
|||||:|||||:|||||:|||||
131 G1yAlaLysAsnGlyLysAsn.AlalysLysGluAspSerAspGluAsp 147
317 AGAAGAAGAAAGCAGACAGAGAGACAACAACAGCAGCAGTGTAGCGC 366
|||:|||||:|||||:|||||
147 LuAspAspAspAspAspAspLysAspSerAspGluAspGluGluAsp 163
164 GluLysAspGluAspGluProValValLysGlyLys...GlnGlyLys 179
417 AGTA.....ATCTCAGTTGAGAGAGACCAAAACGAGAG 451
|||||:|||||:|||||:|||||
179 sValAlaAlaAlaAlaProAlaSerGluAspGluAspGluGluAspG 196
452 ACATGTTGATCAAGAAACGAGCGAGACAGACTGTGCAAGAGTGGC 501
|||||:|||||:|||||:|||||
196 LuGluGluGluGluGluAspGluGluGluGluAspSerGluGlu 212
502 AGGCG 507
|||:|||||
213 GluAla 214

```

seq_name: SwissProt_39:GAG_MSVMO

seq_documentation_block:
ID GAG_MSVMO STANDARD; PRT; 538 AA.

```

AC P03334;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GAG POLYPROTEIN R65 [CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN
DE P12; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10].
GN GAG.
OS Moloney murine sarcoma virus.
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11809;
RN [1]
RP SEQUENCE FROM N.A. (PROVIRUS).

```

RX MEDLINE=62039559; PubMed=6170110;
RA Reddy E.P., Smith M.J., Aaronson S.A.;
RT "Complete nucleotide sequence and organization of the Moloney murine
RT sarcoma virus genome.";
RL Science 214:445-450(1981).

RN [2]
RP SEQUENCE FROM N.A. (CLONE 124 CIRCULAR).
RX MEDLINE=82115347; PubMed=6173134;
RA van Beveren C., van Straaten F., Galleshaw J.A., Verma I.M.;
RT "Nucleotide sequence of the genome of a murine sarcoma virus.";
RL Cell 27:97-108(1981).

RN [3]
RP SEQUENCE OF 2-17, AND MYRISTOYLATION.

RX MEDLINE=83169654; PubMed=6340098;
RA Henderson L.E., Kruttsch H.C., Oroszlan S.;
RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an
RT unusual post-translational protein modification.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).

CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

DR EMBL: J02266; AAA46499.1; -.
DR EMBL: V01185; CAA24507.1; -.
DR PIR: A03932; FOMVM.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF01140; gag_MA; 1.
DR Pfam: PF01141; gag_P12; 1.
DR Pfam: PF02093; Gag_P30; 1.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; ZnF_C2HC; 1.
KW Coat protein; Core protein; Nucleoprotein; Polyprotein; Myristate.
FT CHAIN 2 131 CORE PROTEIN P15.
FT CHAIN 132 215 INNER COAT PROTEIN P12.
FT CHAIN 216 478 CORE SHELL PROTEIN P30.
FT CHAIN 479 534 NUCLEOPROTEIN P10.
FT LIPID 2 2 MYRISTATE.
FT VARIANT 519 519 R -> K (IN CLONE 124).
SQ SEQUENCE 538 AA; 61209 MW; D78326F3B5701E56 CRC64;

```

alignment_scores:
Quality: 91.00 Length: 163
Ratio: 1.071 Gaps: 9
Percent Similarity: 52.147 Percent Identity: 26.380

Alignment_block:
US-09-823-101-4 x GAG_MSVMO ..

Align seg 1/1 to: GAG_MSVMO from: 1 to: 538

```

84 AAAAAAGGAGGGAACAGGAGGACATAGCGAGTGTCCGGG.....CT 127
|||||:|||||:|||||:|||||
353 LysValLysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLe 369
128 GGAAGCTGTGTCACCGCGCTACAAA.....A 153
|||||:|||||:|||||:|||||
369 uGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGlu 386
154 CTCGAAAGAGGAACATTTGTCACGCGCTCCACACTACGCGCACAGACT 203
|||||:|||||:|||||:|||||
386 sProGlyGlnGluThrAsnValSerMetSerPheIleTyrPheIleAla 402
204 .....AGAGAAACACATCTACTA..... 221

```

```

403 PROASPILLEGLYARGLYSLEUGLARGLEUGLNASPLEUARGSNLYSTH 419
222 .....GAACTCAGACATATATATAGTAGACAG 252
419 rLEUGLYASPLEUVALARGLUALAGLUALGILIEPHASN..LYSRYG 435
253 AAAGCCACATAGAGCTACTGCATTAAGTACCACTGCCGAAGACACGT 302
435 LUTHRPROGLUGLU.....ARGLUGLUARG 443
303 ATCAGACACTGACGAGAGAAACGACGACAGAGACAAACCAACGA 352
444 ILAETARGGLUARGGLUGLUGLUGLUGLARGHARGHARGHARGH 460
353 GCCAGCGTCGACGCGATGACGCGACGAGATATCCGGAAGAACGGGG 402
460 U.GLNYSGLUYS...GLUARGASPARGARGARGHISARGLUMETSER 475
403 AAAGACAGATGAAGTAATCTCAGTTGACGAAGACCAAGCAGAGAGA 452
476 ARGLEULEUALATHRVALLVALSER.....GLYGLNARGLNASPAR 489
453 CGATGTTGATCAAGAAACCCAGCGCCAGACAGACTGT 489
489 GGLNGLUGLUGLUGLARGARGARGSERGLNLEUASPCYS 501
seq_name: SwissProt_39:INCE_CHICK

```

```

seq_documentation_block:
ID INCE_CHICK STANDARD: PRT: 877 AA.
AC P53352;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INNER CENTROMERE PROTEIN (INCENP).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RX MEDLINE=94012983; PubMed=8408220;
RA Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.;
RT "Molecular analysis of the INCENPs (inner centromere proteins):
RT separate domains are required for association with microtubules
RT during interphase and with the central spindle during anaphase.";
RL J. Cell Biol. 123:373-385(1993).
CC -1- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY
CC CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC NETWORK DURING INTERPHASE.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: INCENP I AND INCENP II (SHOWN
CC HERE). ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----

```

```

DR EMBL; Z25419; CAA80906.1; -.
KW Cell division; Microtubules; Coiled coil; Centromere; Mitosis;
KW Cell cycle; Nuclear protein; Alternative splicing.
FT DOMAIN 503 733 COILED COIL (POTENTIAL).
FT VARSPPLIC 716 753 MISSING (IN ISOFORM INCENP I).
FT VARIANT 257 257 A -> T.
FT VARIANT 471 471 E -> Q.
SQ SEQUENCE 877 AA; 100940 MW; AFA703149F55352 CRC64;

```

```

alignment_scores:
Quality: 91.00 Length: 179
Ratio: 0.948 Gaps: 5
Percent Similarity: 53.631 Percent Identity: 20.670

```

alignment_block:

US-09-823-101-4 x INCE_CHICK ..

Align seg 1/1 to: INCE_CHICK From: 1 to: 877

```

40 AAACGTAATCTATCAAGAGCAAAAAA.....AAACACC 89
523 ARGLYSGLNYSVALLGUGLUGLUGLUGLUGLUGLUGLUGLUGL 539
90 CGGGGGGAACCCAGGGGACATAGCGAGTCTCCGGGCTGGAACCTGTGT 139
539 SLEUARGARGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 549
140 ACCCGGCTACAAAATCTCAAGAGCAACATTGTCAACGCGCTCCACTA 189
550 .....LEUGLINALARG.....GUARGLALUGLUGLUGLUGLUGLUGL 558
190 CGCGGACACAGCTAGAGAAACACATCTACTAGAACCTCAACATATAT 239
559 LEUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 575
240 ACATAGTACAGAGAACGCCACATAGACTACTCTCGCA.....T 277
575 PGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 592
278 AAAGTACAACTGCGCAAGACAGCTATCAGAGCTGACGACAGACAGAAC 327
592 YLSYSGALALALALALALALALALALALALALALALALALALALALAL 608
328 GAGGAC..... 333
609 GLNASPGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 625
334 ACAGAGACAAACCAACGACGCGTGAACGCGATGAGACAGACGCA 382
625 UATGARGHISLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 642
383 GATACCCGGAAGAACGGGGAAGAACAGATG.....AAA 417
642 TGLALARGLYSLLEALAGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUG 658
418 GTATCTCACTTGAACGACGCAACCAAGCCGAGAGAGAGATTTATCAAG 467
659 GLNLEUALALAGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUG 675
468 AAACGACGCGACAGACAGCTGACAGAGTGCAGAG 504
675 SLYSGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUG 687
seq_name: SwissProt_39:GAG_MLVHO
seq_documentation_block:
ID GAG_MLVHO STANDARD: PRT: 540 AA.
AC P21435; Q80877; Q80878; Q80879; Q80880;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE GAG POLYPROTEIN [CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12;
 DE CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10].
 GN GAG.
 OC HOMULV murine leukemia virus (Mus hortulanus virus).
 OC Viruses: Retrovirdae; Retroviridae; Gammaretrovirus.
 ON NCBI_TaxID=11799;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90051094; Pubmed=2554579;
 RA Voytek P., Kozak C.A.;
 RT "Nucleotide sequence and mode of transmission of the wild mouse
 RT ecotropic virus, Homulv."
 RL Virology 173:58-67(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M26528; AAA5464.1; -
 DR PIR: A32594; FOMVHL.
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_P12.
 DR InterPro: IPR003036; Gag_P30.
 DR InterPro: IPR001878; ZnF_CCHC.
 DR Pfam: PF01140; gag_MA.1.
 DR Pfam: PF01141; gag_P12.1.
 DR Pfam: PF02093; Gag_P30.1.
 DR Pfam: PF00098; zf-CCHC.1.
 DR SMART: SM00343; ZnF_C2HC.1.
 KM Core protein: Coat protein; Nucleoprotein; Polyprotein; Myristate.
 FT CHAIN 2 129 CORE PROTEIN P15.
 FT CHAIN 130 217 INNER COAT PROTEIN P12.
 FT CHAIN 218 480 CORE SHELL PROTEIN P30.
 FT CHAIN 481 540 NUCLEOPROTEIN P10.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 SQ SEQUENCE 540 AA; 61382 MW; 5F25150086F9E95 CRC64;

alignment_scores:
 Quality: 90.50 Length: 165
 Ratio: 1.090 Gaps: 10
 Percent Similarity: 50.303 Percent Identity: 29.091

alignment_block:
 US-09-823-101-4 x GAG_MLVHO ..
 Align seg 1/1 to: GAG_MLVHO from: 1 to: 540

```

84 AAACCCGGGGGAAACGAGGACATAGCCAGTCCCGG.....CT 127
   |||:||||| |||:|||| |||:|||| |||:|||| |||:||||
355 LysValValysGlyIleThrgInglProAsngInSerProSerAlaPhele 371
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
128 GGAACCTGGGATACCCGGCTACAAA.....A 153
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
371 uGluurGleuLysGluAlaIyTArGArTyThProTyAsProGluA 388
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
154 CTCCAAGAGAGAAACATGTTCACGCGCTCCACTACCGGCACACAGCT 203
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
388 sPrroGlyGIngluThrAsnValSerMetSerPheIleTrpInSerAla 404
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
204 .....AGAGAAACATCTACTA..... 221
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
405 ProAspIleGlyArgGlyLysleuGluArgLeuGluAspLeuLysSerLys 421
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
222 .....GAAACTCACACATATATACATAGTACAGAG 252
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
421 rIeuGlyAspLeuValArgGluAlaGluArgIlePheAsn..LysArgG 437
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
253 AAACGCCACATAGAGCTACTCGCATAAAGTACACTGCGCAAGACACGT 302

```

```

437 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
   luThrProGluGlu.....ArgGluGluArg 445
303 ATTCAGACTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
446 lIeaArgArgGluThrgInglLysGluGluArgGluAlaGluAsnGl 462
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
353 GCCACGCTGAGCGCGATGAGCAGACGCGAGATACCCGAAAGAGCGGG 402
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
462 u.GluArgGluLys...GluArgAspArgArgHIsArgGluMetSer 477
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
403 AAAGACAGATGAAGTAATCTCACTTGACGAGACCAAGAGAGAGAGA 452
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
478 LysLeuLeuAlaThrValValSer...GlyGluArgGluAspArgGln 493
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
453 CGATGTTGATCAAGAAACACGCGAGACAGAG...TGTGGA 492
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
493 yGlyGluArgArgArgProGlnLeuAspLyspInGlySala 507
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
seq_name: SwissProt_39:YL02_CAEEL
seq_documentation_block:
ID YL02_CAEEL STANDARD; PRT; 304 AA.
AC P34396;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOPHETICAL 33.9 KDA PROTEIN P10E9.2 IN CHROMOSOME III.
GN P10E9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10986; AAA28013.1; -
DR PIR: S44801; S44801.
DR WormPep: F10E9.2; CE00146.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 33925 MW; 60999E5AE3402A2F CRC64;


alignment_scores:  

  Quality: 90.00 Length: 156  

  Ratio: 1.139 Gaps: 5  

  Percent Similarity: 50.641 Percent Identity: 23.077



alignment_block:


```

US-09-823-101-4 x YLU2_CAEL ..

Align seg 1/1 to: YLU2_CAEL from: 1 to: 304

```

34 GGGGAAAAAGCTAATCTATCATAGAGGAAAAA.....AAAAA 83
138 GilylslelearglyserthrargalathrargylslyserValanly 154
84 AAAACCCGGGGGGAACCCAGGAGACATAGCGGTCCGGCTGGAAC 133
154 sserlyssercllyargthrarghlslyslsvalproaspclnargl 171
134 TGTGTATC.....CCGGCTACAAAAATCCAAAGAGG 165
171 eutrptryphleuylrleuSerlysproValal..... 182
166 AAMCATTTGTCACGCGCTCCACTACCGCGACACACAGCTAGAGAACAT 215
183 .....ValSerleuprotThrAspAlathrTrhHisasple 194
216 CTACTAGAACTCATACATATATACATAGTACAGAGAACGCCACATAG 265
194 uMetAlaIargleuylrThrleuLeuGlYalSerGlyLysThHisMetI 211
266 AGCTACTCGCATAAAGTACAACTGCCGAAGACACGTATCAGAGACTGAC 315
211 lethrleu.....ArgThrclnlysserSerSerGluGlu 223
316 GAGAGAAAGAACGACGACAGAGAGACAAACCAACGCCAGCTGAGCG 365
224 GtlyGlyglubhesercllyarglyasp...ProAlarleuclnValG1 239
366 CGATGACGAGGACGACAGATACCCGAAAGAGCGGAGAAACAGACATGA 415
239 yAspSerlystrArgleuSeraspLeuylsIethrArgHlslyMetI 256
416 AAGTAATCTCAGTTGACGAGAACCAAAAGCGAGAGACGATGTTGATCAA 465
256 euteleuaspValasp...AspGlInlyssaspGlyserGlyGlu 271
466 AGAACCAGCGCGAGACA 483
272 LysLysglulyslysser 277

seq_name: swissprot_39:ATRX_HUMAN

seq_documentation_block:
ID ATRX_HUMAN STANDARD: PRT: 2375 AA.
AC P46100; P51068; Q15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97123494; PubMed=89968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97385582; PubMed=9244431;
RA Villard L., Lossi A.M., Cardoso C., Proud V., Chiaroni P.,
RA Colleaux U., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding

```

```

RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 743 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Stayton C.U., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKCI in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP VARIANTS ATR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Afimos S., Azenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RL Nat. Genet. 17:146-148(1997).
RN [7]
RP VARIANT JM GLN-2014.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Gecz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [8]
RP VARIANT ATR-X LEU-129.
RX MEDLINE=20123062; PubMed=10660327;
RA Fichera M., Romano C., Castiglita L., Falla P., Ruberto C., Amata S.,
RA Greco D., Cardoso C., Fontes M., Ragusa A.;
RT "New mutations in XNP/ATR-X gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RL Hum. Mutat. 12:214-214(1998).
CC -I- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPISING SEVERE
CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL
CC ABNORMALITIES, AND ALPHA-THALASSEMIA.
CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME
CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
CC MICROGENITALISM AND EARLY DEATH.
CC -I- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -I- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```


1364 ArgLysIleLeuLysAspAspLysLeuArgThrGluThrGlnAsnAlaLe 1380
 459 TGATCAAGAAACCGACGCGACAGACAGTGTGCAGAGTGCAGG 504
 : ::::::::::::::: ||||| ||
 1380 uLysGluGluGluArgArgLysArgIleAlaGluArgGluArg 1395

seq_name: SwissProt_39:MST2_DROHY

seq_documentation_block:

ID MST2_DROHY STANDARD; PRT; 1391 AA.

AC 008696;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE AXONEME-ASSOCIATED PROTEIN MST101(2).

GN MST101(2).

OS Drosophila hydei (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7224;

RN [1]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RX MEDLINE=95045538; PubMed=7957199;

RA Neesen J., Padmanabhan S., Buennemann H.;

RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid

RT motif representing the major component of the sperm-tail-specific

RT axoneme-associated protein family Dms101 form extended

RT alpha-helical rods within the extremely elongated spermatozoa of

RT Drosophila hydei."

RL Eur. J. Biochem. 225:1089-1095(1994).

CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTYES AND EARLY

CC SPERMATIDS.

CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.

CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT

CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM

CC REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: X73481; CAA51876.1; -.

DR PIR: S34154; S34154.

DR HSSP: P01032; 1C5A.

DR FlyBase: FBgn0020733; Dhyd\mst101(2).

KW Sperm: Repeat: Multigene family; Polymorphism.

FT DOMAIN 332 1268 59 X 16 AA APPROXIMATE TANDEM REPEATS OF

FT [KRI-K-X-C-X-X-A-K-X-K-X-X-X-E.

SO SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

alignment_scores:

Quality: 88.50 Length: 200

Ratio: 0.903 Gaps: 6

Percent Similarity: 49.000 Percent Identity: 19.500

alignment_block:

US-09-823-101-4 x MST2_DROHY ..

Align seg 1/1 to: MST2_DROHY from: 1 to: 1391

1 CAGGAATTACTAAGATGATGATATCATCATCGAGAAAGCTAATTC 50
 ::::::::::::::: ::::: |||||::: ::
 720 GluGluLeuAlaLysLysIleLysLysAlaAlaGluLysLysCysLys 736
 51 TATCAAGAGCAAAAAAAAAA.....AAAAAAAAAAAAAAAACCCG 91

736 sLysLeuAlaLysLysLysLysAlaGluGluLysAsnLysLeuLysLysG 753
 92 GGGGAAACACGAGGACATACCGAGTGTCCCGGCTGGAAACGTGTGTAC 141
 : ::::: ||| |||
 753 LysLysLysGlyLysLysAla..... 760
 142 CCGGCTCAAAACTCCAAAGAGAAACATTGTCAACGCGCTCCACTACC 191
 : ::::::::::: ||| :::::
 761LeuLysGluLysLysCysArgLysLeuAlaLysLys 773
 192 GCGCACACAGCTAGAGAAACATCTACTAGAAATCACAACATATATAC 241
 : ::::: |||
 773 sLysAlaAlaGluLysLysLysCysLysGluAlaAlaLysLysGluLysG 790
 242 ATAGTACAGAAACCCACATAGAGCTACTCCGATTAATACAACTCGG 291
 : ::::::::::: |||
 790 LuAlaAlaGluLysLysLysCysGlu.....LysThrAla 801
 292 CAAAGACACGTATCAGAGACTGACGAGAGAAAGAACGACAGAGAGA 341
 : ::::::::::: |||
 802 LysLysArgLysGluGluAlaGluLysLysCysGluLysThrAlaLys 818
 342 CAAA.....CCAACGAGCCAGC 358
 : ::::: |||
 818 sLysArgLysGluAlaAlaGluLysLysCysGluLysAlaAlaLysL 835
 359 GTAGAGCGGATGAGCGACGACGATACCCGAAAGAACGGGAAAGA 408
 : ::::::::::: |||
 835 ysArgLysGluGluAlaGluLysLysCysGluLysThrAlaLysLys 851
 409 CAGATGAAGTATCTCAGTTGACGAAGCAAAAGAGAGACGATGT 458
 : ::::: |||
 852 ArgLysGluThrAla.....GluLysLysCysGluLysAlaAlaLys 865
 459 TGATCAAGAAACGAG..... 474
 : ::::: |||
 865 aLysLysArgLysGluAlaAlaGluLysLysCysGluLysAlaAlaLys 882
 475CGGAGACAGACCTGTCCAGAAAGTGCAGAGC 504
 : ::::: |||
 882 ysLysArgLysGluAlaAlaGluLysLysCysAlaGluAlaAlaLys 898

THIS PAGE BLANK (USPTO)


```

136 .....LysGluAspProAspArgGluAlaArgProGlu.Thr..... 147
292 CAAGACACGCTATCAGAGCTGACGAGAGAGAAAGACGACGAGAGAGA 341
148 .....HisLeuAspGluAspGluAspGlyLysLysAspLysArgSerSe 162
342 CAAGACACGAGCGACGCGTGGCGGATGAGCGAGCGAGATATCCCGA 391
162 ArgProArgSerGln...ProArgAspProAlaAlaLysArgArgProL 178
392 AAGAACGGGGGAAAGACAGATGAACTATCTCAGTTGACGAAGACCAA 441
178 ysgGluAlaGluProGluGln.....ValAlaProGluThrProGlu 191
442 AAGCGAGAGACGATGTTGATCAAGAAAGACGCGAGAGACA 483
192 AspArgAspGluAspGluArgGluGluLysArgArgLysThr 205
seq_name: sp_rodent:Q9QXX6

```

```

seq_documentation_block:
ID Q9QXX6 PRELIMINARY: PRT: 1620 AA.
AC Q9QXX6:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
GN DNMT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Margot J.B., Aguirre-Arteta A.M., Di Giacco V.B., Pradhan S.,
RA Roberts R.J., Cardoso M.C., Leonhardt H.;
RA "The genomic organization of the mouse DNA methyltransferase.";
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYL CYTOSINE.
CC -1 SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DR EMBL: AF162282; AAF19352.1; -
DR MGD: MGI:94912; Dnmt1.
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR002857; Znf-CXXC.
DR Pfam: PF01426; BAH; 2.
DR Pfam: PF00145; DNA_methylase; 3.
DR PRINTS: PR00208; zf-CXXC; 1.
DR PRINTS: PR00105; C5METHYTRFAS.
DR SMART: SM00439; BAH; 2.
DR PROSITE: PS00094; C5_MTASE_1; 1.
DR PROSITE: PS00095; C5_MTASE_2; 1.
DR MethyLtransferase; Transferase.
KW SEQUENCE 1620 AA; 183187 MW; 4F9A98CEAF09F037 CRC64;

```

```

alignment_scores:
Quality: 109.50 Length: 148
Ratio: 1.335 Gaps: 7
Percent Similarity: 55.405 Percent Identity: 27.703

```

alignment_block:

US-09-823-101-4 x Q9QXX6 ..

Align seg 1/1 to: Q9QXX6 from: 1 to: 1620

```

54 CAAGACGAGAAAAAAGAAAAAAGAAAAAAGCCGGGAGAAACGAG 103
193 GUGGUARGAspGlnAspLysArgGluValAlaSerPthGluSerGI 209

```

```

104 GGGACATACGAGAGTGTCCCGGCTGAAGTGTGATCCCGGCTAC.... 149
209 yAlaAlaAlaValAlaGluLysLeuGluGluValThrAlaGlyThrGlnL 226
150 .....AAACTCCAAAGAGAAACATTTGACACGCGCTCCACTACC 191
226 euGIProGluGluProGlyGluGluGlnGluAspAsnArgSerLeuArg 242
192 GCGCACACGACTGAGAAACACATCTACTGAAACTCACAACTATTATAC 241
243 ArgHisThr...ArgGluLeuSerLeuArgArgLysSer..... 254
242 ATAGTACAGAGAAACGCCATAGAGCTACTCCATTAAGTACACTGCG 291
255 .....LysGluAspProAspArgGluAlaArgProGlu.Thr..... 266
292 CAAGACACGCTATCAGAGCTGACGAGAGAGAAAGACGACGAGAGAGA 341
267 .....HisLeuAspGluAspGluAspGlyLysLysAspLysArgSerSe 281
342 CAAGACACGAGCGACGCGTGGCGGATGAGCGAGCGGAGATTAACCGA 391
281 ArgProArgSerGln...ProArgAspProAlaAlaLysArgArgProL 297
392 AAGAACGGGGGAAAGACAGATGAAAGTATCTCAGTTGACGAAGACCAA 441
297 ysgGluAlaGluProGluGln.....ValAlaProGluThrProGlu 310
442 AAGCGAGAGACGATGTTGATCAAGAAAGACGCGAGAGACA 483
311 AspArgAspGluAspGluArgGluGluLysArgArgLysThr 324
seq_name: sp_rodent:Q9DBR7

```

```

seq_documentation_block:
ID Q9DBR7 PRELIMINARY: PRT: 1004 AA.
AC Q9DBR7:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 120001SF06RIK PROTEIN.
GN 120001SF06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staib L.F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Rint B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004785; BAB23563.1; -
DR MGD: MGI:1918986; 120001SF06RIK.
DR InterPro: IPR002110; ANK.

```


OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
 RA Mewes H.W., Mayer K., Schueller C.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
 DR EMBL: AL021687; CA16708.1; -;
 DR EMBL: AL161550; CAB7892.1; -;
 DR InterPro: IPR002885; PPR.
 DR InterPro: IPR001878; znf_CCHC.
 DR Pfam: PF01535; PPR; 8.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; znf_C2HC; 1.
 DR Zinc Finger.
 KM
 SQ SEQUENCE 1260 AA; 143517 MW; 626AC4A2F03F02B CRC64;

alignment_scores:
 Quality: 100.50 Length: 176
 Ratio: 1.047 Gaps: 7
 Percent Similarity: 54.545 Percent Identity: 24.432

alignment_block:
 US-09-823-101-4 x 049677 ..

Align seg 1/1 to: 049677 from: 1 to: 1260

```

31 TCAGGAAAAAACGTAATTCATCAAGACGAAAAAAGAAAAA 80
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
360 SearGtLulysArGtLysArGtLysArGtLysLysLysSe 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||
81 AAAAAAACCCGGGGGAAACGAGGACATAGCGAGTCCCGCTGA 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||
376 rLysLysGln.....TyrAspSerAspSerLeuSerPheGluGly.... 389
   |||:|||||:|||||:|||||:|||||:|||||:|||||
131 AACTGTGATACCCGGCTACAAATCCAAAGAGAAACATTTGCAACGC 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||
390 ..SerGtLysArpSerTyrArGtLeuSerArGtArGtHstHrLysHis 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||
181 GCTCCACTACCCGCGCACACAGTA..... 204
   |||:|||||:|||||:|||||:|||||:|||||:|||||
406 ValAsp...ProSerAlaSerLeuLysSerGluValTyrHisGlnGlyAs 421
   |||:|||||:|||||:|||||:|||||:|||||:|||||
205 .....GAGAAACATCTACTAGAAACTCACAACATATATACATA 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||
421 nSerHstArGtLulysGlnHstLysTyrArGtLysGlnLys.....HisG 434
   |||:|||||:|||||:|||||:|||||:|||||:|||||
245 GTTACAGAGAACGCCACATAGACCTACTCGCAATAGTCAATCACTGCCGCA 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||
434 LulysArGtLysGlnLysValAspArGtProSerAlaSerAspAspSer 450
   |||:|||||:|||||:|||||:|||||:|||||:|||||
295 AGACAGCTTACAGACAGTACGACGAGAGAAAGACGACGAGAGACAA 344
   |||:|||||:|||||:|||||:|||||:|||||:|||||
451 AspTyrTyrArGtSerAsnSerSerArGtLysLysArGtSerGlnAspSry 467
   |||:|||||:|||||:|||||:|||||:|||||:|||||
345 ACCAACGACCCAGCGTAGCGC..... 366
   |||:|||||:|||||:|||||:|||||:|||||:|||||
467 rLysSerHstHstArGtLulysGlnValHisSerAsnAspProValS 484
   |||:|||||:|||||:|||||:|||||:|||||:|||||
367 ..GATAGCGACGAGCGCAGATACCCAAAGAAAGACGGGAAAA.....AGA 408
   |||:|||||:|||||:|||||:|||||:|||||:|||||
484 erGtLulysSerGlnLysGlnHstLysTyrSerGlnSerGlnLysLysIleGlnArG 500

```

```

seq_name: sp_invertebrate:Q9W3D1
seq_documentation_block:
ID Q9W3D1 PRELIMINARY; PRT; 1183 AA.
AC Q9W3D1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG12109 PROTEIN (CHROMATIN ASSEMBLY FACTOR-1 P180 SUBUNIT).
OS CAF1-180 OR CG12109.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Feriz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitai M., Katush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mounklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Tyler J., Kadonaga J., Kobayashi R.;
RT "Drosophila CAF-1 subunits."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003444; AAF6399.1; -
DR EMBL: AF367177; AAK31263.1; -

```

DR FlyBase: FBgn0030054; Cafl-180.
SQ SEQUENCE 1183 AA; 133494 MW; 994DB57F6B401B9B CRC64;

alignment_scores:

Quality: 100.00 Length: 140
Ratio: 1.282 Gaps: 6
Percent Similarity: 55.714 Percent Identity: 26.429

alignment_block:
US-09-823-101-4 x Q9W3D1 ..

Align seg 1/1 to: Q9W3D1 from: 1 to: 1183

```
78 AAAAAAAAAACCGGGGGAACCAACGAGACATAGCGAGTCCCGGCT 127
   |||:::|||||::: |||::: |||
466 GlnylsgluthrAlrgLeuasnAlrgLysSerleuProgluValThrGly.. 481
128 GGAACACTGTGTACCCGCGCTACAAACTCCAAAGAGAAACATTGTCAA 177
   |||::: |||::: |||::: |||::: |||
482 .....AlaprolysglLeuThrProLysglIngluAlrgLeuMetG 495
178 CGCGCTCACACCGCGCACACACGTAGAGAAACATCTTCTGAACACT 227
   |||::: |||::: |||::: |||
495 luGlnAlrgLysAlrgLgluLgluLysgluLysLeuAlgluLglu 511
228 CACAACATATATACATAGTACAGAAACGCCCATAGAGCTACTCGCAT 277
512 ArgtAlrgLeuLys.....GlnGlnAspLysgluLhsAlrggluLglu.. 525
278 AAGTACAACTGCGCAAGACACGATATCAGACTGACGAGAGAGAAAC 327
   |||::: |||::: |||::: |||::: |||
526 .....LysLysgluLgluAlrgAspLgluLysgluLgluLglu 537
328 GACACAGAGAGACAAACCAACGACCGAGCT.....GAGCGCGCA 368
   |||::: |||::: |||::: |||::: |||
537 rglYsLeuGluAlrgAspLgluLysglu..GlnGlnAlrgLysMetGluLysgl 553
369 TGAGGACAGACGCGAGATACCGCAAGAAAGAGGAAACGACGATGAAG 418
553 ||| ||||| ||||| ||||| |||||
553 glulYsgluAlrg.....LysAlrggluAlgluLglu 563
419 TAATCTGAGTTGACGAGACCAACCAAGCAGAGAGATGTGATCAAGA 468
   ||| ||| |||::: |||::: |||::: |||::: |||
563 alAspSerLysAsnGluLgluLysAlrgLysAlrgAsnGluAlgluLysglu 579
469 AACCAAGCGCGACAGACAGAC 486
   |||||::: |||
580 ValGlnAlrgLysLysAsp 585
```

seq_name: sp_bacteria:087306

seq_documentation_block:

ID 087306 PRELIMINARY; PRT; 406 AA.
AC 087306;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE ELPB2 (ORF29/ELPB2).
GN ORF29/ELPB2.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=297;
RA Akins D.R., Bourrell K., Calmano M.J., Hagman K., Norgard M.V.,
RA Radolf J.D.;
RT "An new animal model for studying Lyme disease spirochetes in the
RT host-adapted state";
RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC PLASMID=CP18-2;
RX MEDLINE=20143776; PubMed=10678977;
RA Calmano M.J., Yang X., Popova T.G., Clawson M.L., Akins D.R.,
RA Norgard M.V., Radolf J.D.;
RT "Molecular and evolutionary characterization of the cp32/18 family of
RT supercoiled plasmids in Borrelia burgdorferi 297.";
RL Infect. Immun. 68:1574-1586(2000).
DR EMBL; AF023853; AAC34964.1; -;
DR EMBL; AF169008; AAF29801.1; -;
DR InterPro: IPR003483; OsPEF.
DR Pfam: PF02471; OsPEF; 1.
KW Plasmid.
SQ SEQUENCE 406 AA; 46543 MW; A3D92D3F42D3EB88 CRC64;

alignment_scores:

Quality: 99.00 Length: 140
Ratio: 1.207 Gaps: 5
Percent Similarity: 58.571 Percent Identity: 23.571

alignment_block:
US-09-823-101-4 x 087306 ..

Align seg 1/1 to: 087306 from: 1 to: 406

```
60 AAAAAAAAAAAAAAAAAAAAAAAAAACCGGGGGAACCAACGAGGACA 109
   |||:::|||||::: |||::: |||::: |||::: |||
142 GluGluLysgluLysgluLysgluLysAlgluLysAlrgLysglY.. 157
110 TAGCGAGTGTCCCGCGCTGAAACTGTGTACCGCGCTACAAACTCCAA 159
   |||::: |||::: |||::: |||
158 .....LysgluValIleGluLgluLgluLysLysglu 168
160 AAGAGGAACATTGTCAACGCGCTCCACTACCGCGCACACAGCTAGAGAA 209
   |||::: |||::: |||::: |||::: |||
168 lngluLgluThrAlaLysLysAla.....LysAlaGlnLysgluLyl 181
210 ACACATCTACTAGAAACTCACACATATATACATAGTACAGAAACGCC 259
   |||::: |||::: |||::: |||
181 S.....ArgGluAlrg 185
260 ACATAGACTACTCGCATATAGTACAACTGCGCAAGACAGCTATCAGAG 309
   |||::: |||::: |||::: |||::: |||
185 lngluLysIleGlnGlnLysgluLgluLgluLgluLgluAlrgAlaLysglu 201
310 ACTGACGAGAGAGAAACGACGACGAGAGAGACAAACCAACGACCGACG 359
   |||::: |||::: |||::: |||
202 GluGluLgluLgluLgluLgluAlrgAlaLysgluLgluLgluLgluAlrgAr 218
360 TGAGCGGATGAGCGCAGAG...CGCAGATACCCGCAAGACGGGGAAGA 406
   |||::: |||::: |||::: |||
218 GluLysgluLgluLgluLgluLgluLgluAlrgAlaLysgluLgluLgluAl 235
407 GACAGATGAAGTAACTCACTGACGAGAGACCAACCAACGAGAACAGAT 456
   |||::: |||::: |||::: |||::: |||
235 rglGlnAlrgAlrgAlaLysgluLgluLgluLgluLgluLgluAlrgAlaLysglu 251
457 GTTGATCAAGAAACCGAG 474
   |||::: |||
252 GluGluLgluLysAlrgGln 257
```

seq_name: sp_plant:Q9LH98

seq_documentation_block:

ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T1998.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBITaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002057; BAB03174.1; -;
SQ SEQUENCE 2081 AA: 232852 MW: D3603E1F85EFFF29 CRC64;

alignment_scores:

Quality:	99.00	Length:	160
Ratio:	1.151	Gaps:	5
Percent Similarity:	53.750	Percent Identity:	21.875

alignment_block:

US-09-823-101-4 x Q9LH98 ..

Align seg 1/1 to: Q9LH98 from: 1 to: 2081

```
37 AAAAAACGTAATCTATCAAGAGAAAAA.....AAAAA 86
|||||.....|
1213 LysLysGlnThrSerValGluGluAsnLysGlnLysGlnThrLysLys 1229
87 ACCCGGGGGAACCAAGCGGAC.....ATAGCAGTGTCCCG 124
|.....|
1229 sGluLysAsnLysProLysAspLysLysAsnThrThrLysGlnSerG 1246
125 GCTGAACACTGTGTACCCGCTACCAAACTCCAAAAGAGAAACATTGT 174
|||.....|
1246 LysLysLysLysGlnSerMetGlnSerLysGlnLysGlnLysGln 1262
175 CAACGCGCTCCACTACCGCGCAGACAGCTAGAGAAACACATCTAGAA 224
|||.....|
1263 GlnLysSer.....G 1266
225 ACACAGCAACATATATACATAGTACAGAAACCCACATAGACTACTCG 274
|.....|
1266 AlaThrThrGlnAlaAspSerAspLysLysAsnGlnLysLeuMetG 1283
275 CATTAAGTCAACTCGCAAAAGACAC.....GTATCAGAGCTGAC 315
|.....|
1283 LalaAspSerGlnAlaAspSerHisSerAspSerGlnAlaAspSerAsp 1299
316 GAGAGAGAAAGAC.....GACAGAGAGACAAACCAAGACCA 356
|||.....|
1300 GlnSerLysAsnGlnLysLeuMetGlnAlaAspSerGlnAlaThrThrG 1316
357 GCGTGAGCGGATGAGGACAGCGCAGATACCCGAAAGAAACGGGAAAA 406
|||||.....|
1316 naTgaAsnAsnGlnLysAspTrpLys..... 1324
407 GACAGATGAAGTAATCTCAGTTCAGCAAGACCAAAAGCAGAGACGAT 456
|||.....|
1325 .....LysGlnThrSerValAlaGluAsnLysLysGlnLysGlnThr 1338
457 GTTGATCAAAAGAACCAAGCGCAGAGACAG 486
|.....|
1339 LysGlnGlnLysAsnLysProLysAspAsp 1348
```

seq_name: sp_invertebrate:Q23319

seq_documentation_block:

```
ID Q23319 PRELIMINARY; PRT; 389 AA.
AC Q23319;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE ZC43.4 PROTEIN.
GN ZC43.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBITaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z75553; CAA99946.1; -;
SQ SEQUENCE 389 AA: 44984 MW: 5812D07D30329297 CRC64;
```

alignment_scores:

Quality:	98.50	Length:	154
Ratio:	1.187	Gaps:	5
Percent Similarity:	53.896	Percent Identity:	24.026

alignment_block:

US-09-823-101-4 x Q23319 ..

Align seg 1/1 to: Q23319 from: 1 to: 389

```
58 AGCAAAAAAAAAAAAAAAAAAAACCGGGGGAACCAAGCGGA 107
|||||.....|
162 ArgAlaLysSerArgSerGlnLysAsnValProAlaTrpSerPheAsp 178
108 CATAGCGAGTGTCCCGGC.....TGGAAC 133
|.....|
178 ProAlaLeuLeuProGlnGlnLysProLeuProGlnAlaThrLysV 195
134 TGTGTACCCGGCTACAAACTCCAAAGAGAAACATTGTCAAGCGCT 183
|.....|
195 alTyrTrpAsnAlaValLysSerValGlnAlaProThrProGlnGlnAla 211
184 CCACCTACCGCGCACACAGCTAGAGAAACACATCTACTAGAACTCACAAC 233
|||||.....|
212 ProValAlaGlnAspValAlaValAsnGlnLeu..... 222
234 ATATATACATAGTACAGAGAAACGCCACATAGACTACTCGCATAAAGTA 283
|||.....|
223 .....SerGlnValLysLysTrpGlnLysLeu..... 232
284 CAATCGCGCAAGACACGATATCAGAGACTGACAGAGAAAGCAAGCAG 333
|.....|
233 .....LysAlaGlnMetGlnGlnLysArgTrpGlnGlnAsp 246
334 AGAGAGAGCAAAACCAAGCAGCGCTGAGCGGAGAGAGAGAGCGCAG 383
|||.....|
247 .....AspValAlaAsnLysLeuAsnGlnValValGlnLysLeuLys 261
```



```

384 ATATCCGGAAGAAACGGGAGAAACACGATGAAGAATCTCAGTTGACG 433
      |||||:::|||||  |||
261 rlrhralysaspsergilyssserlin.....LysG 272
      |||||:::|||||  |||
434 AAGACCAAAAGCAGAGAAGCAGATGTTGATCAAGAAACGCGAGACA 483
      |||||:::|||||  |||:::  |||||
272 laspluugiugluugiugluspaspaspaspaspasppluugiuglu 288
      ::::  ::::|||
484 GACTGTGCAGAA 495
      ::::  ::::|||
289 gluasnvaiglu 292

seq_name: sp_virus:055708

seq_documentation_block:
ID      055708      PRELIMINARY:      PRT:      456 AA.
AC      055708:
DT      01-JUN-1998 (TREMblrel. 06, Created)
DT      01-JUN-1998 (TREMblrel. 06, last sequence update)
DT      01-JUN-1998 (TREMblrel. 06, last annotation update)
DE      HYPOTHEtical. 52.8 KDA PROTEIN.
OS      Chilo iridescent virus (CIV) (Insect iridescent virus type 6).
OC      Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus
CX      NCBI_TaxId=10488;
CC      (1)
      SEQUENCE FROM N.A.
      RA      Bahr U., Tidon C.A., Darai G.;
      RL      Virus Genes 0:0-0(1997).
      DR      EMBL; AF003534; AAB94419.1; -.
      KW      Hypothetical protein.
      SQ      SEQUENCE 456 AA; 52915 MW; 290CD897A855B798 CRC64;

```

[illegible]

seq_name: sp_plant:Q9SY74

```
seq_documentation_block:
ID      Q9SY74      PRELIMINARY;      PRT;      836 AA
```

DT 01-MAY-2000 (TremBrel. 13, Created)
DT 01-MAY-2000 (TremBrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBrel. 17, Last annotation update)
DE F14N23.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxId=3702;
RN [1]
RP
SEQUENCE FROM N.A.
RA Shinn P., Dunn P., Walker M., Buehler E., Kim C., Alatafi H.,
RA Atrajeno R., Conn L., Conway A.B., Gonzalez A., Hansen N.F., Huizar L.,
RA Kremeretskata I., Lenz C., Li J., Liu S., Lucero S., Rowley D.,
RA Schwartz J., Toriumi M., Vystotskaia V., Yu G., Davis R.W.,
RA Pederspiel N.A., Theologis A., Ecker J.R.
RA Genomic sequence for Arabidopsis thaliana BAC F14N23 from Chromosome
1.";
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005489; AAD32882.1; -;
DR InterPro; IPR000571; 2f-CCCH.
DR Pfam; PF00642; 2f-CCCH.2;
DR SMART; SM00356; Znf_C3H1; 2.
QO SEQUENCE 836 AA; 93171 MW; 87F5C291FCE46AEA CRC64;

```
alignment_scores:
  Quality: 98.00
  Ratio: 1.089
  Percent Similarity: 50.279
  Length: 179
  Gaps: 10
  Percent Identity: 25.698
```

alignment_block:

US-09-823-101-4 X Q9SY74

Align seg 1/1 to: Q9SY74 from: 1 to: 836

54 CAAGAGGAAAAAAAAAAAAAAAAAAAAAACC GGGGGGAAC CAG 103

[illegible]

441 AAAG.....CGAGAGACGATGTTG 460

799 sLysGluArgArgHisArgHisArgLysArgArgArgT^HArgLysAsnSerA 816

461 ATCAAGAACCACGCCGAGACAGACTGTGCAGAA 495

```

||:::|| |::||::|| |::||
816 enaenc|naenbro|wec|u|serc|u|va|c|u 827

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-1 to: PIR-68: * out_format : pfs

Date: Jan 17, 2002 4:03 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ .n2p.model -DEV=xlp
-O:/cgr2.1/USPTO.spool/US09823101/rnatat_17012002.155137_22030/app_query.fasta_1.2239
-DB-PIR-68 -OPMT-fastan -SUFFIX=n2p.rpr -GAPOP=12.000
-GAPOP=4.500 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-bioscore2
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MILEN=0 -MAXLEN=200000000
-USER=US09823101.ecgn1_1_200 -NCP=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-1
Query length: 542
Database: PIR_68: *
Database sequences: 219241
Database length: 76174552
Search time (sec): 120.350000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:A41812	- 89.00	170.81	0.0523	305	peroxisome assembly factor-1, F
pir2:C81923	- 86.50	170.51	0.0987	168	probable membrane protein NMA07
pir2:T21776	- 82.00	155.02	0.3503	345	hypothetical protein F3552.1 -
pir2:A36886	- 80.50	156.28	0.3067	203	surface protein P9 negative re
pir2:D83867	- 80.50	149.52	0.5358	457	hypothetical protein BH1740 [Im
pir2:A45989	- 80.00	151.86	0.5964	304	peroxisome assembly factor-1 -
pir2:C84914	- 79.50	150.83	0.6826	303	hypothetical protein At2g47360
pir2:S46187	- 79.50	148.97	0.6932	379	probable membrane protein YK121
pir2:S69731	- 79.50	144.17	0.7212	674	hypothetical protein YDR452w -
pir2:T21884	- 79.00	137.03	0.8681	1398	hypothetical protein F36H1.2 -
pir2:D81387	- 78.50	146.24	0.9131	408	probable integral membrane prot
pir2:S64837	- 78.50	144.47	0.9266	505	hypothetical protein YLR015w -
pir2:S63322	- 77.50	144.71	1.119	381	probable membrane protein YML33
pir2:S67813	- 76.50	142.55	1.56	383	probable membrane protein YDL24
pir2:B85181	- 76.00	144.73	1.74	260	FCA alpha and beta protein [Im
pir2:B82929	- 76.00	144.47	1.74	268	hypothetical protein U0131 [Im
pir2:T18214	- 75.00	129.06	2.55	1323	ATP binding cassette protein -
pir2:A83886	- 74.50	139.71	2.65	325	hypothetical protein BH1889 [Im
pir2:S57191	- 74.00	138.34	2.68	383	hypothetical protein YK1219w hc
pir2:S36954	- 74.00	139.66	3.01	288	cytochrome-c oxidase (EC 1.9.3.
pir2:L53362	- 74.00	138.18	3.02	305	peroxisome assembly factor-1 -
pir2:T24908	- 74.00	131.86	3.21	734	hypothetical protein T1407.2 -
pir2:T15863	- 73.00	123.03	3.40	1667	hypothetical protein C3E6.1 -
pir2:S53782	- 73.00	137.07	3.96	305	peroxisome assembly factor-1 -
pir2:D64060	- 73.00	134.69	4.04	406	plus assembly protein p1lg hom
pir2:T18440	- 73.00	114.55	4.77	4550	hypothetical protein C0425w -
pir2:B71801	- 72.50	135.24	4.56	335	probable undecaprenyl-phosphate
pir2:T25338	- 72.50	122.28	4.87	870	hypothetical protein ZK1037.11
pir2:S72489	- 72.00	133.78	4.99	171	hypothetical protein T199.2 -
pir2:T25625	- 72.00	134.94	5.19	306	hypothetical protein C44C3.4 -
pir2:F86337	- 72.00	133.83	5.20	310	hypothetical protein yjcd [Im
pir2:T23187	- 72.00	133.07	5.27	383	hypothetical protein K01D12.4 -
pir2:C84949	- 72.00	130.76	5.37	505	NADH dehydrogenase (ubiquinone
pir2:C71607	- 72.00	122.56	5.75	1351	hypothetical protein P80745w -
pir2:S48974	- 71.50	142.33	5.54	111	hypothetical protein YHR130c -
pir2:T31694	- 71.50	133.11	5.88	264	hypothetical protein D1065.2 -
pir2:T28973	- 71.50	134.68	5.90	278	gene 278 protein - phase Pfs
pir2:T21020	- 71.50	132.58	6.01	360	hypothetical protein F1603.6 -
pir2:T04508	- 71.50	132.03	6.03	380	hypothetical protein F8F16.230
pir2:S48920	- 71.50	132.06	6.03	381	hypothetical protein YHL048w -

pir2:S78176 - 71.50 129.27 6.17 532 i transcription initiation fac
pir2:T00092 - 71.50 125.84 6.35 803 i hypothetical protein 7 - Str
pir2:T01220 - 71.00 143.90 6.21 81 i hypothetical 9k protein - Pse
pir2:S07729 - 71.00 138.44 6.49 156 i hypothetical protein 4 - Pat
pir2:S36955 - 71.00 133.33 6.77 288 i cytochrome-c oxidase (EC 1.9

seq_name: pir2:A41812

seq_documentation_block:

peroxisome assembly factor-1, PAF-1 - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 10-Dec-1999
C:Accession: A41812
R:Shimozawa, N.; Tsukamoto, T.; Suzuki, Y.; Orit, T.; Shiryoshi, Y.; Mori, T.; Fujik
Science 255, 1132-1134, 1992
A:Title: A human gene responsible for Zellweger syndrome that affects peroxisome asse
A:Reference number: A41812; MUID:92188187
A:Accession: A41812
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-305 <SH1>
A:Cross-references: GB:M6852; NID:g189848; PIDN:AAC12785.1; PID:g189849
A:Note: sequence extracted from NCBI backbone (NCBIP:87610)
C:Superfamily: peroxisome assembly factor-1
C:Keywords: transmembrane protein

alignment_scores:

Quality: 89.00 Length: 171
Ratio: 1.011 Gaps: 8
Percent Similarity: 51.462 Percent Identity: 23.977

alignment_block:

US-09-823-101-1/rev x A41812 ..
Align seg 1/1 to: A41812 from: 1 to: 305

```

478 GGTAGTCTTTCTTTTGAATGATCTGTTGATGACATGATGATTTCTGAA 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 G1G1YLeu11Leu18nheLeu11eLeu18nG1YLeu18nLeu18n 165
428 CTTGSCACAG.....GGAATTAACCAATTTTGGACGACAA 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 Leu11nG1YLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 181
387 AATTCTTAAGATATTCTGCTGCTCTTATTAATAATTTGGAATATTAATA 338
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 Ina1n1e1nG1YLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 195
337 GAACAAAGATGACGATATGATGACGATGACGATGACGATGACGAT 288
|||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....Leu11nG1YLeu18nLeu18nLeu18nLeu18nLeu18n 203
287 TTTTGCATATTTCTGCTGCTCTTCTGCTCTTCTGCTCTTCTGCT 256
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 e.Leu11eLeu18nLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 219
255 .....TGCATATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 Leu18nSe11nG1YLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 Leu18nLeu18nLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 244
167 TCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 .....AlaLeu18nG1YLeu18nLeu18nLeu18nLeu18nLeu18nLeu18n 258
CTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 68
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 CysG1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n1n 273

```



```

156 CTGCGCTCTCTCTGCGCTACCTTCCACAGCAATGACCTCTCTTTTCT 107
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
164 .....ValThrThrIlePheIle 169
106 CTTTTCCTATTGTTCTACGCCCTCCCTCACCCTAAC..... 68
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
170 ILeYrAlaPheLeuAsnTyPheProPheIgluAsnAsnLysAsnAspVa 186
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
67 .....CTTAGAGCTCTGCTGCTGTTACCATATGATGAGCTTCA 28
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
166 lCysTyThrIleuArgPheSerAsnGlyMetGluLeuValArgIlePheS 203
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
27 CTCAGATC 20
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
203 eTThrIleu 205
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
seq_name: p1r2:A36886
seq_documentation_block:
surface protein P4g negative regulator par - Streptococcus sobrinus (strain MT3791)
C:Species: Streptococcus sobrinus
C>Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Jun-2000
C:Accession: A36886
R:Takahashi, I.; Okahashi, N.; Hamada, S.
J. Bacteriol. 175, 4345-4353, 1993
A:Title: Molecular characterization of a negative regulator of Streptococcus sobrinus su
A:Reference number: A36886; MUID:93323312
A:Accession: A36886
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <TAK>
A:Cross-references: GB:D13323; NID:g217037; PIDN:BA02581.1; PID:g425488
A:Note: It is uncertain whether Met-1 or Met-7 is the initiator
C:Genetics:
A:Gene: par
A:Start codon: GTG
C:Superfamily: Streptococcus sobrinus surface protein P4g negative regulator par
C:Keywords: DNA binding

```

```

alignment_scores:
Quality: 80.50 Length: 181
Ratio: 1.134 Gaps: 10
Percent Similarity: 39.227 Percent Identity: 25.414

```

```

alignment_block:
US-09-823-101-1/rev x A36886 ..

```

```

Align seg 1/1 to: A36886 from: 1 to: 203

```

```

532 TGTTCCTCCCTCTTTTCTTTT..... 509
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
33 CysIleProPhePhePheSerIleuSerArgLeuThrIleG1 49
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
508 .....TTTTTTTTTTTGG..... TTTTTTAACCTT 481
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
49 yIleSerAsnIlePhePheAsnTyTrpIleuProPhePheAlaProP 66
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
480 GAGGTAGTTTCTTTAGTATCTGTTCATATGCGCATATTTCG 431
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
66 heAlaLeuAlaLeuLeuValIle..... HisGluGlyIle..... 78
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
430 AACCTGGCAGCAGGAATTAAAAACCAATTTTTGTAGTCAGAAATTCCT 381
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
78 ..... 78
380 AAAAGTATTCTGTCTCTTATATAAATTGGAATATATAAAGAACAA 331
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
78 ..... 78
330 GAATCAGCATATTCACACATCAGACAAATCAGTGCATATTTCG 281
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
79 .....HisGlyPheTyPheIle 84
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

280 ATATT...CCTTCCTGCTCTTTTCTCTGCAATATAGAAAACGTGAGAC 234
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
85 LeuPheLysProGluAsnProLeuLys...TyrGlyThrAspTrpArgLe 100
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
233 TGGATTTCAGTGGCAGTGAAGCATGAGTGAACCTACACAGAAATGTCCT 184
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
100 uGlyLeuPheAsnAlaThrSerProGlySerArgTyProArg..... 114
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
183 GTTTCATATTAAAGTCATGTCGTGCTGCTGCTGCTGCTGCTGCTGCT 134
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
115 .....SerGluMetLeuIleIleTyIleu 122
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
133 TCCCA...CTAATGACCTCTCTTTTCTCTGCAATATTGCT..... 91
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
123 AlaProPheValIleuThrSerIleuLeuThrIleuLeuAlaIleuG 139
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
90 .....CTTAGCCCTTCCCTCAGTAACTTACCTAGAGTCTG 55
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
139 lYrThrLeuSerProLeuAlaTyIleuPheLeuAlaValIle 152
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
seq_name: p1r2:D83867
seq_documentation_block:
hypothetical protein BHI740 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: D83867
R:Nakami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: D83867
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <SHO>
A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BA05459.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BHI740

```

```

alignment_scores:
Quality: 80.50 Length: 195
Ratio: 0.885 Gaps: 9
Percent Similarity: 46.667 Percent Identity: 24.615

```

```

alignment_block:
US-09-823-101-1/rev x D83867 ..

```

```

Align seg 1/1 to: D83867 from: 1 to: 457

```

```

505 TTTTGTGTTTGTGTTTAACTGAGTAGTGTCTTCTTTAGTAT 456
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
67 PheCysPheTrpIleGluGlyAlaSerGlySer.....ThrIle 79
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
455 CTTGTTCATTTAGCATATTTCTGAACCTGGCAGGAAATTAATAAAC 406
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
79 uValPheSerPheGlyGlnAsnLeuAsnIleIleSer..... 91
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
405 CAATTTTGTAGTCAGAAATTCCTTAAAGATATTTCTGTGCTTATATA 356
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
92 ..MetPheLeuPheIleProPheIleSerThrPheLeuSerSerValGly 107
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
355 AATTGGAAATATAAAGAA.....CAAGAATCAGGCATATTCAC 312
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
108 TyrIleAspSerMetLysHisTrpValGlnLysGlnGluLysArgAr 124
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
311 ACTCAGACAAATCAGTGCATATTTCGCAATATTTC..... 273
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
124 gluArgProAspHisThr..AlaPheTrpLeuThrSerIleValGlyLe 140
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
272 TTCCGTGCTCTT..... 261
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

140 uleuLeuAsnPhcglYSerLeuAlaIleValLysArgIleIleGluGlyS 157
    ::::::::::::::
260 ..TTCTGTGATATAGAAAAGCTGAGAGCTGATTTTTCAGTGACATGA 213
    |||||:::|||||:
157 erPheserSerPhearGlu..... 163
212 GCATGAGTGGACCTACACAGATGTTCTGTTTCCATATTAAAGTCATG 163
    ::::::::::::::
164 .....GlnArgLeuMetLeuValIleLeuArgGlyPh 174
162 TGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
    ||| ||||| ||| ::||| ::|||
174 eglySerCysLeuLeuTrpSer...ProPhemeValAsnIleGlyLeu 190
122 .....GACCTCTCTTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCT 87
    |||||:::|||||:
190 leuThrIlePhesAspLeuSerTrpPtyrGlnIleGlyGlyPhGlyLeu 206
86 GCCCTTCCCTCCTCACTTAACCTTAGAGCTGCTGCTGCTGCTGCTGCT 37
    :: ||| ||||| ||| |||
207 leuMetAlaLeuIleTyrllePhelLeuPhetrlleuPhetrlleuProAlaI 223
36 GAGCTCTCAGTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3
    ::::: ||| ::::: |||
223 eglnPhesAspHisAspProValIleGluHisHis 234

seq_name: p1r2:A45989

```

```

seq_documentation_block:
  peroxisome assembly factor-1 - Chinese hamster
  C:Species: Cricetus griseus (Chinese hamster)
  C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Jun-2000
  C:Accession: A45989; A53782; S31621
  R:Thieringer, R.; Raetz, C.R.H.
  J. Biol. Chem. 268, 12631-12636, 1993
  A:Title: Peroxisome-deficient Chinese hamster ovary cells with point mutations in peroxi
  A:Reference number: A45989; MUID:93286102
  A:Accession: A45989
  A:Molecule type: mRNA
  A:Residues: 1-304 <THI>
  A:Cross-references: GB:217220; NID:949474; PIDN:CAAT8929.1; PID:949475
  R:Tsukamoto, T.; Shimozawa, N.; Fujiki, Y.
  Mol. Cell. Biol. 14, 5458-5465, 1994
  A:Title: Peroxisome assembly factor 1: nonsense mutation in a peroxisome-deficient Chine
  A:Reference number: A53782; MUID:94309666
  A:Accession: A53782
  A:Molecule type: mRNA
  A:Residues: 1-304 <TSU>
  A:Cross-references: GB:D30618; NID:9531027; PIDN:BAA06308.1; PID:91109617
  C:Superfamily: peroxisome assembly factor-1
  C:Keywords: transmembrane protein

```

```

alignment_scores:
  Quality: 80.00      Length: 171
  Ratio: 0.930       Gaps: 8
  Percent Similarity: 50.292   Percent Identity: 22.807

```

alignment_block:

US-09-823-101-1/rev x A45989 ..

Align seg 1/1 to: A45989 from: 1 to: 304

```

478 GGTAGTTTCTTCTTTAGTATCTGTTCCATTAATGCGATATTCTGAA 429
    ||:::|||||:
148 GtGluLeuLeuAsnPhcglYSerLeuAlaIleValLysArgIleIleGlu 164
428 CTTGGCAGAG.....GGAATTAATAAACCAATTTTCTGATGACGA 388
    |||||:::|||||:
164 leuThrIlePhesAspLeuSerTrpPtyrGlnIleGlyGlyPhGlyLeu 180
387 AATTTTAATAAGATTTTCTGCTCTTATTAATAATTTGGAATATATAAAA 338

```

```

180 InAsnIleArgGluValGlyPhesPtyrMetAsnArgGluLeu..... 194
    :::::
337 GAACAAGAAATCAGCACAATTAATTCACACACACTGACAAAATCAGTCGACTA 288
    |||||:::|||||:
195 .....LeuTrpHisGlyPheAlaGluPh 202
287 TTTTGGCATATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
    :||| :::::|||||:
202 e.LeuIlePhelLeuLeuProleuLeuAsnIleGlnIlyPheValAlaLys 218
255 .....TCATATATAGAAAAGCTGAGAGCTGATTTTTCAGTGGC 218
    ||||| ||||| ||||| ::::: |||
219 leuSerSerTrpCysIleProleuThrGlyAlaAlaSerSerAspSerAl 235
217 ACATGAGATGATGAGACCTACACAAAGATGTTGCTGTTTCCATATTAAAG 168
    |||||:::|||||:
235 alaValaSerSerGly.....LysGluCys..... 243
167 TCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
    ::||| ::::: ||||| |||||
244 .....AlaLeuCysGlyGluTrpProThrMetProHisThrIleGly 257
117 CTCCTTTTCTCTTTTCTGATTTTGTCTAGCCCTTCCCTCCTTAAC 68
    ::||| ::||| ::|||
258 CysGluHisValPheCysTyTrpCysValLysSerSer.....PheLe 272
67 CTTAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18
    ::::: ||| |||||
272 uphAspMetTrpPhetrlleuPhetrlleuPhetrlleuPhetrlleu 289
17 TGGTTCCAATG 7
289 alGlnProleu 292

seq_name: p1r2:C84914

```

seq_documentation_block:

```

  hypothetical protein At2g47360 [imported] - Arabidopsis thaliana
  C:Species: Arabidopsis thaliana (mouse-ear cress)
  C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
  C:Accession: C84914
  R:Lin, X.; Kaul, S.; Roundsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
  M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
  euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
  Nature 402, 761-766, 1999
  A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
  A:Reference number: A84420; MUID:20083487
  A:Accession: C84914
  A>Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-303 <STO>
  A:Cross-references: GB:AE02093; NID:92275214; PIDN:AAB63836.1; GSPDB:GN00139
  C:Genetics:
  A:Gene: At2g47360
  A:Map position: 2

```

```

alignment_scores:
  Quality: 79.50      Length: 91
  Ratio: 1.728       Gaps: 3
  Percent Similarity: 50.549   Percent Identity: 28.571

```

alignment_block:

US-09-823-101-1/rev x C84914 ..

Align seg 1/1 to: C84914 from: 1 to: 303

```

348 AATATATAAAGAAAGAAATCAGCACAATTAATTCACACACTGACAAAAT 299
    |||||:::|||||:
7 LysTrpArgGlyGluAlaGluGlnLeuValValLysProPhetrlleuVal 23
298 CACTGTGACTATTTTGGCAATATTTCTTCTGCTCTTTTCTGCTGATAT 249

```


alignment_block:

US-09-823-101-1 x S69731 ..

Align seg 1/1 to: S69731 from: 1 to: 674

```

256 GAGAAAAAGCAGACGAGAA.....ATATGCCAAATAGTCACAG 296
|||||.....|
463 GtulsylsAsplysLysLysProIlethrArgLysgluLeu1 479
297 TGATTTGTCGAGTGTGTAATATGCGTATCTTGTCTTTTATAT 346
:::|||||:::
479 egLuarGTYSerLleValasnIleGlySerValIleProThrPhea 496
|||||.....|
347 TTCCAAAATTT.....TATAAGACACAGAAATACTT..... 378
|||||.....|
496 snProSerPheArgIleTrpGluTYrAsnIleThrAspIleValasnAsp 512
379 .....TTAGAAATTCCTGACTACAAAAAATTTG.....TTTTTTAATT 416
:::|||||:::
513 SerAsnPheAlaValSerGluTYrLysPro.TTPAspGluPheGluS 529
|||||.....|
417 CCCTGTGCCAAGTTTCAGAAATATGCCATATGACACAGAT..... 457
|||||.....|
529 eteLusnLysIleMetGluAspSerLeuLeuGluAspGluMetAspSer 545
|||||.....|
458 ...AACTAAAGAGAAAGAACTACCTCAAGGTTAAAAACCAAAAAAAA 504
|||:::|
546 SerAsnIleGluValGlyIleAsnArgGluLysMetGlyGluLysLysAs 562
505 AAAAAAAAAAAAAAAAAAGGGGGGGAACA 532
562 nLysLysLysLysLysAsnAspLysThr 571

```

seq_name: plr2:T21884

seq_documentation_block:

```

hypothetical protein F36H1.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21884
R:Kershaw, J.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19482
A:Accession: T21884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1398 <NTL>
A:Cross-references: EMBL:Z68760; PIDN:CAA92996.1; GSPDB:GN00022; CESP:F36H1.2
A:Experimental source: clone F36H1
C:Genetics:
A:Gene: CESP:F36H1.2
A:Map position: 4
A:Introns: 38/3; 60/3; 94/3; 152/3; 191/3; 257/2; 290/3; 323/3; 379/1; 439/1; 485/2; 552

```

alignment_scores:

Quality:	79.00	Length:	63
Ratio:	2.026	Gaps:	2
Percent Similarity:	61.905	Percent Identity:	33.333

alignment_block:

US-09-823-101-1/rev x T21884 ..

Align seg 1/1 to: T21884 from: 1 to: 1398

```

519 TTTTNTTTTNTTTTNTTTTGGTTTAACTTGAGTACTTTT 470
|||||.....|
491 PhePheAlaIlePheLeuPheGlyMetPheSerLeuThrPheThrPhe 507
|||||.....|
469 CTTCCTTAGTATCTGTCCATATGCAATATTTCAACTTGACCA 420
:::|||||:::
507 lleuIleAlaIleSerAsnSerValThrAlaTYrLeuIleSerTPSerV 524

```

```

419 GCGAATTAAAAACCAATTTTGTACTCAGAAATCTTAAAGTATTTC 370
|||.....|
524 alPheLeu.....LeuIlePheIleIle.....Phe 532
369 TGTGCTCTTATAAAATTTGGAATATATAAGACACAA 331
|||||.....|
533 CysSerLeuIleValValIaTYrTYrGlyAspArgLys 545

```

seq_name: plr2:D81387

seq_documentation_block:

```

probable Integral membrane protein Cj0430 (Imported) - Campylobacter jejuni (strain N
C:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: D81387
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chli
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; MUID:20150912
A:Accession: D81387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CA874266.1; PID:9696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0430

```

alignment_scores:

Quality:	78.50	Length:	119
Ratio:	1.454	Gaps:	6
Percent Similarity:	45.378	Percent Identity:	26.050

alignment_block:

US-09-823-101-1/rev x D81387 ..

Align seg 1/1 to: D81387 from: 1 to: 408

```

541 TTTGCCCCCTTTGTTCCCCCTTTTNTTTTNTTTTNTTTT..... 497
|||||.....|
212 PheAlaIalCysPheSerProLeuValPheValTYrPhePheTYrThr11 228
496 .....TGGTTTTTTTAAAC 484
228 eTYrArgLeuThrPheGlnLysTYrLysAsnLeuLeuThrPhe..... 242
483 CTTGAGTAGTTTCTTCTTTTATCTGTTCCATTAAGCATATTTT 434
|||||.....|
243 .....LeuMetSerValThrPheValPheCysLeuLeu 253
433 CTGAACTTGGCACAGGAATTAANAACA.....AT 402
|||||.....|
254 LeuSerLeuArgGlnLysLeuPheLeuAspAspPheLeuProPheCysVa 270
401 TTTTGTAGCAGAAATCTTAAAGTATTTCTGCTCTTATAAA..... 356
|||||.....|
270 IileCysThrProLeuLeuIleLysThrLeuMetGlnSerTYrArgValA 287
355 .....AATTGGAATATATAAGAACACAAAGATACACCAT 320
:::|||||:::
287 rgLeuLeuValPheArgLeu.ArgTYrLysIlePheIleGluCysSer11 303
319 AATTCAACCACTCAGACAAATATCATCTGTACTATTTTGGCATATTTCTTC 270
|||||.....|
303 eIle.....PheLeuIlePheCysTYrPheLeu1 313
269 CTGTC 265
|||
313 leVal 314

```

seq_name: plr2:S64837


```

372 AATACCTTAAGAAATTTCTGACTACAAAATGGTTTAAATTCCTG 421
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
133 eutyrgluasnlysalatrpasnthrlystyrphepheasnalamet 149
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
422 ...TGCCAA...GTTCAGAATATGCCATATGACACAGATTAATAAA 465
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
150 ValysgngluAlaIleuAlaIleuThrLeuGluProPheSerLeu 166

466 GAAGAAACTACTCAAGGTTAAACCAAAA 499
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
166 slyAspGluAlaIleuValLysSerPheLys 177

seq_name: pir2:s67813

seq_documentation_block:
Probable membrane protein YDL248w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D0215
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 20-Jun-2000
C:Accession: S67813
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.; Wagner, G.
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S67655
A:Accession: S67813
A:Molecule type: DNA
A:Residues: 1-383 <RIE>
A:Cross-references: EMBL:Z74296; NID:g1431423; PIDN:CA98828.1; PID:g1431424; MIPS:YDL24
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:CO57
A:Cross-references: SGD:S0002407; MIPS:YDL248w
A:Map position: 4L
C:Superfamily: conserved hypothetical protein YKL219w
C:Keywords: transmembrane protein
E:74-90/Domain: transmembrane #status predicted <TM1>
E:233-249/Domain: transmembrane #status predicted <TM2>
E:256-272/Domain: transmembrane #status predicted <TM3>

alignment_scores:
Quality: 76.50 Length: 128
Ratio: 1.109 Gaps: 6
Percent Similarity: 53.906 Percent Identity: 22.656

alignment_block:
US-09-823-101-1 x S67813 ..

Align seg 1/1 to: S67813 from: 1 to: 383

164 ATGACCTTAATATGAAACAGAACATTTGTGTAGTCCACTCATGCT 213
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
59 LeuSerValTrpTrpLysLeuSerAsnAsnCysIleTyrProLeuIleVa 75
214 CAGTGCACACTGAAATAATCCAGCTCTCCAGTTTCTAT..... 250
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
75 IserLeuIleuValLeuPheLeuGlyProIlePheValIleuValIleCysg 92
251 ..ATGCGAGAAAAGACAGACAGAAAGAA.....TATGCCAAA 266
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
92 IyLeuSerAlaGlyLysSerLeuSerLysGluIleGlnPheCysLys 108
287 ATATGACACTGATTTTG.....TCTGAGTGGTGAATTAAT 321
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
109 GluIleThrGluAsnThrProSerSerAspProHisAspTrpGluValVa 125
322 GGCCTGATCTTTGTCTTTTATATTTCCAAATTTTATTAAGACACAGA 371
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
125 IValAlaAsnIleu.....AsnSerTyrL 133
372 AATACCTTAAGAAATTTCTGACTACAAAATGGTTTAAATTCCTG 421
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
133 eutyrgluasnlysalatrpasnthrlystyrphepheasnalamet 149

```

```

422 ...TGCCAA...GTTCAGAATATGCCATATGACACAGATTAATAAA 465
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
150 AspCysGluLysMetPheArgThrThrValLeuIleuProPheSerLeu 166
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
466 GAAGAAACTACTCAAGGTTAAACCAAAA 499
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
166 slyAspLysAlaIleuValLysSerPheLys 177

seq_name: pir2:B85181

seq_documentation_block:
FCA alpha and beta protein [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: B85181
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: B85181
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <STO>
A:Cross-references: GB:NC_001268; NID:g5302787; PIDN:CA846036.1; GSPDB:GN00140
A:Genetics:
A:Gene: d14182c
A:Map position: 4

alignment_scores:
Quality: 76.00 Length: 121
Ratio: 1.188 Gaps: 4
Percent Similarity: 52.893 Percent Identity: 24.793

alignment_block:
US-09-823-101-1 x B85181 ..

Align seg 1/1 to: B85181 from: 1 to: 260

32 GACTCACATTAATGTGTAACAGACACCACTCTTAAGTTAAGTAGGCA 81
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
77 GluSerAspTyrSerValArgProThrThrProVal..... 89
82 AGGCTAGACAAATATGAGAAAGAAAGAAAGAGAGTCAATAGTGGG 131
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
90 .....GlnGlnProLeuSerGlnLysArgGlyTyrProIleS 103
132 GAAAGTAGGCCAGAGACAGGACGAGACCAATGACCTTAATATGAGAA 181
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
103 eAspHisGlySerPheThrGlyThrAsp...ValSerAspArgSerSer 118
182 ACAGAACATTTCTGTAGTCCACATGCTCAGTCCCACTGAAATAT 231
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
119 ThrValLysLeuPheValGlySerValProArgThrAlaThrGluGlu 135
232 CAGCTCTCCAGTTTC.....TATA 251
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
135 uIleArgProTyrPheGluGlnHisGlyAsnValIleuGluValAlaIleu 152
252 TGCAGAGAAAAGACAGAGAGAAATATGCCAAATATGACACAGATTA 301
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
152 IeLysAspLysArgThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 168
302 TTGTCTGAGTGGTGAATTAATGCTGATTTCTTTGTTCTTTTATATTC 351
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
169 LeuGlyAsn.....SerArgGluLeuTyrPheLeuLysPheG 181
352 AATTTTATAAGA 364
    |||  ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
181 nValSerLeuLys 185

```


THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-1 to: SwissProt_39:* out_format: pfs
Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgnr2.1/USPTO.spool/US09823101/runat.17012002.155138.22061/app-query.fasta.1.2239
-DB=SwissProt_39 -OPMT=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USRR=US09823101@cgnr2.1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPX
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-1
Query length: 542
Database: SwissProt_39:*
Database sequences: 100059
Database length: 3664827
Search time (sec): 66.470000

score_list:	Strd	Orig	ZScore	EScore	len	Documentation
Sequence						
SwissProt_39:PEX2_HUMAN	-	89.00	162.30	0.0750	305	p28328 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	82.00	150.26	0.4336	247	p28328 nephroneolins olivace
SwissProt_39:PEX2_CRIGR	-	80.00	144.64	0.7251	304	p06438 cricetus griseus (d
SwissProt_39:YH16_MNEPOL	-	79.50	141.85	0.8312	379	p38363 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	78.50	137.54	1.0811	505	p43132 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	77.50	137.88	1.3778	381	p53822 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	76.50	135.87	1.7717	383	p00778 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	76.00	137.80	1.9817	268	p00778 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	75.00	122.79	2.7413	1323	p53706 candida albicans (ye
SwissProt_39:YH16_MNEPOL	-	74.50	131.94	2.9338	383	p47187 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	74.00	132.81	3.29305	305	p55098 mus musculus (mouse)
SwissProt_39:YH16_MNEPOL	-	74.00	132.86	3.43714	714	p57058 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	73.00	130.85	4.24305	305	p24392 rattus norvegicus (rat)
SwissProt_39:YH16_MNEPOL	-	73.00	128.51	4.29406	406	p44651 haemophilus influenza
SwissProt_39:YH16_MNEPOL	-	72.50	126.04	4.91487	487	p15374 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	72.00	124.76	5.58505	505	p57263 buchnera aphidicola
SwissProt_39:YH16_MNEPOL	-	71.50	136.16	5.89111	111	p38834 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	71.50	128.66	6.16278	278	p03627 bacteriophage pf3. 32
SwissProt_39:YH16_MNEPOL	-	71.50	127.84	6.19307	307	p09690 mus musculus (mouse)
SwissProt_39:YH16_MNEPOL	-	71.00	126.08	6.25381	381	p38723 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	71.00	137.75	6.5881	81	p25755 pseudomonas putida. by
SwissProt_39:YH16_MNEPOL	-	71.00	132.39	6.79156	156	p15605 paramecium tetraureli
SwissProt_39:YH16_MNEPOL	-	70.00	120.53	9.26524	524	p04870 xenopus laevis (afric
SwissProt_39:YH16_MNEPOL	-	69.50	131.72	9.79118	118	p04522 fasciola hepatica (li
SwissProt_39:YH16_MNEPOL	-	69.50	122.15	10.35381	381	p53344 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	69.50	115.29	10.77882	882	p13342 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	68.50	119.00	13.41437	437	p26641 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	68.50	117.30	13.54542	542	p00976 schistosoma haematoph
SwissProt_39:YH16_MNEPOL	-	68.50	116.81	13.58576	576	p43250 homo sapiens (human)
SwissProt_39:YH16_MNEPOL	-	68.50	114.15	13.80797	797	p00511 neisseria gonorrhoeae
SwissProt_39:YH16_MNEPOL	-	68.50	114.14	13.80798	798	p00514 neisseria meningitidis
SwissProt_39:YH16_MNEPOL	-	68.00	113.83	13.82829	829	p46954 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	68.00	118.66	15.15407	407	p49707 gallus gallus (chicken
SwissProt_39:YH16_MNEPOL	-	68.00	113.16	15.65798	798	p08766 neisseria flavescens.
SwissProt_39:YH16_MNEPOL	-	68.00	110.92	15.861049	1049	p25571 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	67.50	117.20	17.54623	623	p46030 candida albicans (yea
SwissProt_39:YH16_MNEPOL	-	67.50	110.67	18.231386	1386	p06574 marchantia polymorpha
SwissProt_39:YH16_MNEPOL	-	67.00	117.04	19.46390	390	p57558 buchnera aphidicola
SwissProt_39:YH16_MNEPOL	-	67.00	115.95	19.59446	446	p54070 saccharomyces cerevis
SwissProt_39:YH16_MNEPOL	-	67.00	115.73	19.61458	458	p05208 pseudomonas pseudoalid

SwissProt_39:PEP_DROME + 67.00 112.08 20.04 716 p1073 drosophila melanoga
SwissProt_39:CHSA_EMENTI - 67.00 109.24 20.37 1013 p30584 emericella nidula
SwissProt_39:CSA_YEAST + 66.50 116.29 22.05 379 p43542 saccharomyces cere
SwissProt_39:HUNK_MOUSE + 66.50 111.12 22.72 714 p08866 mus musculus (mous
SwissProt_39:KUTL_CAEEL + 66.50 110.78 22.77 744 p02381 caenorhabditis ele

seq_name: SwissProt_39:PEX2_HUMAN

seq_documentation_block:

ID PEX2_HUMAN STANDARD: PRT: 305 AA.
AC P28328:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXISOMAL MEMBRANE
DE PROTEIN 3) (35 KDA PEROXISOMAL MEMBRANE PROTEIN).
GN PMP3 OR PEX2 OR PAF1 OR PMP35 OR PMP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
OX [1]
-PEP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=92188187; PubMed=1546315;
RA Shimozawa N., Tsukamoto T., Suzuki Y., Orit T., Shirayoshi Y.,
Mori T., Fujiki Y.;
RT "A human gene responsible for Zellweger syndrome that affects
peroxisome assembly";
RL Science 255:1132-1134(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CH3CA ZINC FINGER.
RX MEDLINE=93050151; PubMed=1426230;
RA Pataarca R., Fletcher M.A.;
RT "Ring finger in the peroxisome assembly factor-1";
RL FEBS Lett. 312:1-2(1992).
RN [4]
RP VARIANT IRD LYS-55.
RA Shimozawa N., Imamura A., Zhang Z., Suzuki Y., Orit T., Tsukamoto T.,
Osumi T., Fujiki Y., Wenders R.J.A., Beasley G., Kondo N.;
RT "Defective paf gene products correlate with the protein import,
biochemical abnormalities, and phenotypic heterogeneity in peroxisome
biogenesis disorders";
RL J. Med. Genet. 36:779-781(1999).
CC -1- FUNCTION: SOMEWHAT IMPLICATED IN THE BIOGENESIS OF PEROXISOMES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- DISEASE: DEFECTS IN PMP3 ARE THE CAUSE OF ZELLWEGER SYNDROME-1
(ZMS-1), AN AUTOSOMAL RECESSIVE DISORDER DUE TO DEFECTIVE IMPORT
MECHANISMS FOR PEROXISOMAL MATRIX ENZYMES.
CC -1- DISEASE: DEFECTS IN PMP3 ARE THE CAUSE OF INFANTILE REFSUM
DISEASE (IRD). FEATURES INCLUDE EARLY ONSET, MENTAL RETARDATION,
MINOR FACIAL DYSMORPHISM, RETINITIS PIGMENTOSA, SENSORINEURAL
HEARING DEFICIT, HEPATOMEGALY, OSTEOPOROSIS, FAILURE TO THRIVE,
AND HYPOCHOLESTEROLEMIA. THE BIOCHEMICAL ABNORMALITIES INCLUDE
ACCUMULATION OF PHYTANIC ACID, VERY LONG CHAIN FATTY ACIDS
(VLCFA), DI- AND TRIHYDROXYCHOLESTANOIC ACID AND PIPECOLIC ACID.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M86652; AAC12785.1; -;
DR EMBL; M85038; AAA60141.1; -;

DR PIR; A41812; A41812.
 DR MIM; 170993; -.
 DR MIM; 214100; -.
 DR MIM; 266510; -.
 DR InterPro; IPR001841; ZnF_ring.
 DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
 DR Transmembrane; Peroxisome; Zinc-finger; Zellweger syndrome;
 KM disease mutation.
 FT TRANSMEM 140 159 POTENTIAL.
 FT TRANSMEM 195 213 POTENTIAL.
 FT ZN_FING 244 283 RING-TYPE.
 FT VARIANT 55 55 E -> K (IN IRD).
 FT /FTID=VAR_011389.
 SQ SEQUENCE 305 AA; 34896 MW; F62F492D613F2CA CRC64;

alignment_scores:
 Quality: 89.00 Length: 171
 Ratio: 1.011 Gaps: 8
 Percent Similarity: 51.462 Percent Identity: 23.977

alignment_block:

US-09-823-101-1/rev x PEX2_HUMAN ..

Align seg 1/1 to: PEX2_HUMAN from: 1 to: 305

```

478 GGTGGTTTCTCTTTAGTATCTGTCCATATGCGCATTTTCGAA 429
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
149 GLyGlyLeuIleAsnPhelLeuGlnArgGlyLysPheAlaTh 165
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
428 CTGGGCACAG.....GCAATTAACAAACAAATTTTGTAGTCAGA 388
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
165 rLeuThrGlnArgLeuLeuGlyIleHisSer...ValPheCysLysProG 181
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
387 AATTCCTTAAGATATTTCTGTCTGTATTAATAATTTGGAATATTAATA 338
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
181 InAsnIleArgGlyValGlyPheGlyIleMetAsnArgGlyLeu..... 195
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
337 GAACAAAGAAATCAGCCATTAATTCACACACTGACGAAATTCCTGACTA 288
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....LeuTrpHisGlyPheAlaGlyLys 203
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
287 TTTTGGCATATTTCTCTCTGTCTGTCTTTTCTC..... 256
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
203 e.LeuIlePheLeuLeuProLeuIleAsnValGlnLysLeuLysAlaLys 219
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
255 .....TGCATATAGAAAACAGAGAGCTGATTTTTCAGTGGC 218
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
220 LeuSerSerTrpCysIleProLeuThrGlyAlaProAsnSerAspAsnTh 236
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
217 ACTGAGCATGAGTGACCTACACAGAAATGTCGTTTTCATATTAAAG 168
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
236 rLeuAlaThrSerGly.....LysGlyLys..... 244
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TCATGTGCTCTGCTGCTCTGTGCGCTACCTTTCCCACTAATGACCT 118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
245 .....AlaLeuCysGlyGlyTrpProThrMetProHisThrIleGly 258
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
117 CTCTTTTCTCTTTTCTCTATTTTGTGTAGAGCCCTTCCCTCACTTAAC 68
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
259 CysGlnHisIlePheCysTyrPheCysAlaLysSer.....PheLe 273
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
67 CTTAGAGAGCTGTGCTGTGTACCATATGATGACTTTCACCTGACATCTC 18
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
273 uPheAspValTyrPheThrCysProLysCysGlyThrGlnValHisSerL 290
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 TGGTTCCAATG 7
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
290 euGlnProLeu 293
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: SwissProt_39: YMI6_NEPOL

seq_documentation_block:
 ID YMI6_NEPOL STANDARD; PRT; 247 AA.
 AC 09PC94;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN YMI6.
 GN YMI6.
 OS Nephroselmis olivacea.
 OG Mitochondrion.
 OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;
 OC Chlorodendraceae; Chlorodendraceae; Nephroselmis.
 OX NCBI_TaxID=11312;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIES-484;
 RX MEDLINE=99418884; PubMed=10488238;
 RA Turmel M., Lemieux C., Burger G., Lang B.F., Otis C., Plante I.,
 RA Gray M.W.;
 RT "The complete mitochondrial DNA sequences of Nephroselmis olivacea and
 RT Pedinomonas minor: two radically different evolutionary patterns
 RT within the green algae.";
 RL Plant Cell 11:1717-1729(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE TATC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF110138; AAF03203.1; -.
 DR InterPro; IPR002033; UPF0032.
 DR Pfam; PF00902; UPF0032; 1.
 DR PROSITE; PS01218; TATC; FALSE_NEG.
 KM Mitochondrion; Hypothetical protein; Transmembrane.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 SQ SEQUENCE 247 AA; 30505 MW; 3805C1E626C04B35 CRC64;

alignment_scores:
 Quality: 82.00 Length: 196
 Ratio: 0.943 Gaps: 12
 Percent Similarity: 44.388 Percent Identity: 25.510

alignment_block:

US-09-823-101-1/rev x YMI6_NEPOL ..

Align seg 1/1 to: YMI6_NEPOL from: 1 to: 247

```

522 CCCTTTTCTTTTCTTTTCTTTT..... 502
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
82 ProGlyrPhePheYrGlnPheTrpAlaPhePheIleProSerThrTyrGly 98
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
501 .....TTTTGGTTTCTTTTAACTGAGGTAG 474
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
98 nPheGlnArgLeuGlnLeuArgPheLeuSerPhePhePheThrLeuL 115
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
473 TTTTCTCTTTTGTATTTCTTTGTCATTTAGGATTTTCTGAACCTGG 424
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
115 euPhePhe.....SerCysIleIleIle..... 122
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
423 CACAGGAAATTAACAAACAAATTTTGTAGTCAGAAATTTTAAAGTA 374
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

123 .....Ty 123
373 TTTCTGCTCTTATATAATTTGGAAATAT.....AAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| |||
123 rheiIeIleuProgluIleIerPheSerPheleuHsPhegluLysL 140
338 AGAAC.....AAGAAATCAGCCATAATTCACACACACAGAC 304
    ||| ||| ||| ||| ||| ||| ||| ||| |||
140 ySerIeTyIerPheAsnLeuGlnLeuAlaIarGlleSerSerTyIle 156
303 AAAATCAGCTGACATATTTGGCATATTTCCCTCGCTCTTCTCTG 254
157 GlnPheIrnPheGlnIlePheSerTyIerPhe...ValIeuPheGlnCy 172
253 CATATGAAAACTGGAGAGCTGATTTTTCAGTGGCAGCTAGC..... 211
    ||| ||| ||| ||| ||| ||| ||| ||| |||
172 sProIeuPheIrn.....HisPheSerIeuAsnLeuLeuI 186
210 .....ATGAGTGGACCTACACAGAAATGTTCTGTTTTCATAT 172
    ||| ||| ||| ||| ||| ||| ||| ||| |||
186 hrIleSerPheLeu.ValAsnSerIarGlyTyIerTyPheLeuPhe 202
171 AAGGTATGTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
202 v.....IleuAlaIalPheLeuSerProProAspI 213
121 ACCCTCT...TTTTCTCTTTTCTTATTTGTTCTACCCCTCCCTC 75
    ||| ||| ||| ||| ||| ||| ||| ||| |||
213 IeIeuSerIrnPhePheLeuPheSerIeuIleValPheMetTyrgluLeu 229
74 ACTTAACCTTAGAGTCTGCTGCTGTTGATACATAT 39
230 Cys.....ValPheTySerCysPheTyAsp 238

```

seq_name: SwissProt_39:PEX2_CRIGR

```

seq_documentation_block:
ID PEX2_CRIGR STANDARD; PRT; 304 AA.
AC Q06438;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
DE PROTEIN 3).
GN PMP3 OR PEX2 OR PAF1 OR PMP35.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=93286102; PubMed=7685346;
RA Thieringer R., Raetz C.R.H.;
RT "Peroxisome-deficient Chinese hamster ovary cells with point
RT mutations in peroxisome assembly factor-1.";
RL J. Biol. Chem. 268:12631-12636(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94309666; PubMed=8035823;
RA Tsukamoto T., Shimozawa N., Fujiki Y.;
RT "Peroxisome assembly factor 1: nonsense mutation in a peroxisome-
RT deficient Chinese hamster ovary cell mutant and deletion analysis.";
RL Mol. Cell. Biol. 14:5458-5465(1994).
CC -1- FUNCTION: SOMEWHAT IMPLICATED IN THE BIOGENESIS OF PEROXISOMES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z17720; CAA78929.1; -.
DR EMBL: D30618; BAA06308.1; -.
DR PIR: S31621; S31621.
DR PIR: A45989; A45989.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW Transmembrane; Peroxisome; Zinc-finger.
FT TRANSMEM 139 158 BY SIMILARITY.
FT ZN_FING 194 212 BY SIMILARITY.
FT FT 243 282 RING-TYPE.
FT MUTAGEN 246 246 C->Y: LOSS OF FUNCTION.
SO SEQUENCE 304 AA; 34794 MW; 84EC5FA613C148BF CRC64;

```

alignment_scores: Quality: 80.00 Length: 171

Ratio: 0.930 Gaps: 8

Percent Similarity: 50.292 Percent Identity: 22.807

alignment_block:

US-09-823-101-1/rev x PEX2_CRIGR ..

Align seg 1/1 to: PEX2_CRIGR from: 1 to: 304

```

478 GGAGTATTTCTCTTTAGTATCTGTTGATATGACATATTTCTGAA 429
    ||| ||| ||| ||| ||| ||| ||| ||| |||
148 GlyIuLeuIleAsnPheLeuIlePheGlnIysGlyLysPheAlaIrh 164
428 CTTGGCAGAG.....GGAATTAACCAATTTTGTAGTCACA 388
    ||| ||| ||| ||| ||| ||| ||| ||| |||
164 rLeuIrnGluIarGleuLeuGlyIleHisSer...ValPheCysLysProG 180
387 AATTTCTTAAGATATTTCTGCTCTTATATAATTTGGAATATTAATA 338
    ||| ||| ||| ||| ||| ||| ||| ||| |||
180 IAsnIleIarGluIuValGlyPheAspTyIeMetAsnIarGluLeu... 194
337 GAACAAAGATCAGCCATATTCACACACACAGAAATCAGTGCATTA 288
    ||| ||| ||| ||| ||| ||| ||| ||| |||
195 .....LeuIrnHisGlyPheAlaIarGlu 202
287 TTTTGGCATATTTCTCTCTGCTGCTTTTCTG..... 256
    ||| ||| ||| ||| ||| ||| ||| ||| |||
202 e.LeuIlePheLeuLeuProLeuIleAsnIleGlnLysPheLysAlaLys 218
255 .....TGCATATAGAAACTGGAGAGCTGATTTTTCAGTGGC 218
    ||| ||| ||| ||| ||| ||| ||| ||| |||
219 LeuSerSerTrpCysIleIerProLeuIrnGlyAlaIaIaSerIarSperal 235
217 ACTGACATGATGAGACATACACAGAAATGTTCTGTTTTCATATTAAG 168
    ||| ||| ||| ||| ||| ||| ||| ||| |||
235 AleuAlaIaSerGly.....LysGlnCys..... 243
167 TCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
    ||| ||| ||| ||| ||| ||| ||| ||| |||
244 .....AlaLeuCysGlyGluTrpProIleIrnHisIleGly 257
117 CTTCTTTTCTCTCTTTTCTTATTTTCTTACGCCCTCCSCATTAAC 68
    ||| ||| ||| ||| ||| ||| ||| ||| |||
258 CysGlnHisValPheCysTyIerCysValLysSerSer.....Phe 272
67 CTTAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18
    ||| ||| ||| ||| ||| ||| ||| ||| |||
272 uPheAspMetTyrgPheIrnCysProLysCysGlyIleGluValHisSerV 289
17 TGTTTCCATG 7
    ||| ||| ||| ||| ||| ||| ||| ||| |||
289 aIcIrnProLeu 292

```

seq_name: SwissProt_39: COS2_YEAST

```

seq_documentation_block:
ID COS2_YEAST STANDARD: PRT; 379 AA.
AC P38363;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COS2/COS3 PROTEIN.
OS (COS2 OR YBR302C OR YBR212I) AND (COS3 OR YML132W OR YM4987.03).
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE OF 1-47 FROM N.A. (COS2).
RC STRAIN=S288C;
RA Feldmann H., Mannhaupt G., Schwarzlose C., Vetter I.;
RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 30-379 FROM N.A. (COS2).
RC STRAIN=S288C;
RA Feuerhann M., Potier S., Souciet J.-L.;
RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (COS3).
RC STRAIN=S288C / AB972;
RA Bowman S., Louis E.J., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (Jul-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DUF/COS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z36171; CAA85267.1; -
DR EMBL: Z50178; CAA90551.1; -
DR PIR: S46187; S46187.
DR SGD: S0000506; COS2.
DR InterPro: IPR001142; DUF.
DR Pfam: PF00674; DUF; 2.
KW Transmembrane; Multigene family.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
SQ SEQUENCE 379 AA; 45165 MW; 2AAB51B6C103AF61 CRC64;

```

alignment_scores:

Quality:	79.50	Length:	128
Ratio:	1.187	Gaps:	6
Percent Similarity:	52.344	Percent Identity:	23.438

alignment_block:

US-09-823-101-1 x COS2_YEAST ..

Align seg 1/1 to: COS2_YEAST from: 1 to: 379

```

164 ATGACCTATATAGAAACAGAAACATCTGTGAGTGCCTCACTCATGCT 213
      :::::  |||||  ::::::::::::::: |||||:::
59 LeuSerValTrpTrpIysLeuSerAsnAsnGlySileTrpProLeuIleVa 75
214 CAGTGGCACTGAAAAAATCCAGCTCTCCAGTTTCTAT..... 250
      :|||  ::|||:::|
75 IserLeuValLeuPheLeuGlyProIlePheValLeuValIleCysG 92
251 ..ATGCAGAGAAAAAGACAGAGAGAA.....TATGCCAAA 286

```

```

92 IlyeSerArgIysArgSerLeuSerLeuGlnLeuIleGlnPheCysIys 108
      :::::|||||::: |||
287 ATAGTCACAGTCGATTTG.....TCTGAGTGTGCAATTAT 321
      ::|||
109 GluIleThrGluAsnThrProSerSerAspProHisAspTrpGluValVa 125
322 GCGTCAATCTTTGTTCTTTTATATTCACAAATTTTATAGACACAGA 371
      :||| ::|||
125 IAlAlAsnLeuAsnSerIleYrIleGluAsnAsnValTrpAsnThrL 142
372 AATACCTTTAAGAAATTTCTGACTACAAAAAATGCTTTTAAATCCCTG 421
      |||||
142 YsIYrIpe.....PhePheAsnAlaMet 149
422 ..TGCCAA...GTTCAGAAATATGCCATATGACACAGTAACTAAAA 465
      |||||
150 ValCysGlnGluAlaPheArgThrThrLeuLeuGluProPheSerLeuL 166
466 GAGAGAAACTACCTCAAGGTTAAACCAAA 499
      |||
166 slyAspIysAlaIAlaIysValIysSerPheIys 177

```

seq_name: SwissProt_39: YLIR_YEAST

```

seq_documentation_block:
ID YLIR_YEAST STANDARD: PRT; 505 AA.
AC P43132;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 58.3 KDA PROTEIN IN PPRI-SNF7 INTERGENIC REGION (ORFX).
YLR015W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Vandenbol M., Portetelle D., Hliger F.;
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-472 FROM N.A.
RC STRAIN=S288C;
RA MEDLINE=95179524; PubMed=7874498;
RA Levin D.E., Stevenson W.D., Watanabe M.;
RT "Evidence against the existence of the purported Saccharomyces
RT cerevisiae PKC2 gene.";
RL Curr. Biol. 4:990-995(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z73187; CAA97537.1; -
DR EMBL: L34405; AAA34835.1; -
DR SGD: S0004005; YLR015W.
DR InterPro: IPR003877; SPRY.
DR SMART: SM00449; SPRY; 1.
KW Hypothetical protein.
SQ SEQUENCE 505 AA; 58347 MW; DSB33221E9F10379 CRC64;

```

alignment_scores:

Quality:	78.50	Length:	174
Ratio:	0.882	Gaps:	7
Percent Similarity:	51.149	Percent Identity:	24.138

alignment_block:

US-09-823-101-1 x YLIR_YEAST ..

US-09-823-101-1 x YLR_YEAST ..

Align seg 1/1 to: YLR_YEAST from: 1 to: 505

```

80 GAAGGCGCTAGACAAATAGAAAAAGAGAAAGAGAGCTATTAGNG 129
   ||| |||:||||: |||:||||: |||: |||: |||:
275 GIUASPALAGLUNSGIYProserArgLysLeuLysGInGluThrTh 291
   :|||: |||:||||: |||:||||: |||: |||: |||:
130 GGGAAAGGTAGGCCAGAGAGCA..... 151
   :|||: |||:||||: |||:||||: |||: |||: |||:
291 rsnLysSGIuPhGInArgAlaLeuLeuLunSplEGluTyraAspV 308
   :|||: |||:||||: |||:||||: |||: |||: |||:
152 ..GGCAGAGACCAATGACCTTAATATGGAAGAACAGACATTCTGTGA 199
   |||:||||: |||:||||: |||:||||: |||: |||: |||:
308 aIValArgAspGInIleAlaIleArgTyLysAsnGIn..... 320
   |||:||||: |||:||||: |||:||||: |||: |||: |||:
200 GGTCCACTCATGCTCAGTGCACCTGAA.....AATCCAGCTCTCCACT 243
   |||:||||: |||:||||: |||:||||: |||: |||: |||:
321 .....LeuPhGInArgAlaThrAspTyValLysThrThLysProG 335
   :|||: |||:||||: |||:||||: |||: |||: |||:
244 TTTCATATGACAGAGAAAGACAGAGAAATATGCCAAATAGTCA 293
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
335 uTyTyTyrSerAspLysArgGluArgGInAspTyTyTyGIn..... 349
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
294 CAGTGATTTGTCTGAGTGTGAATTATGCTGATTCCTTTGTTTGA 343
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
350 .....LeuGluAspSer..... 353
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
344 TATTTCCAATTTTATTAGACGACGAAATACTTTAAGAATTTCTGAC 393
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
354 TyLeuAlaIlePhGInAsnGlyLysTyLeuGlyLysAlaPhGInL 370
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
394 TACAAAAAATGCTT.....TTTAATCCCTGTCACCAAGTTCAGAAAT 437
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
370 nLeuLysProLeuLeuProPhoSerGInTyraSngInLysP 387
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
438 ATGCCATA.....ATGAAACAGATTAATAAAGAG 469
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
387 heTyTyLeuGlyTyTyTrGInHisGlyLysAlaArgAspGInSerAsn 403
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
470 AAACACTACTCAAGTTAAAAAACCAAAAAA..... 519
   |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
404 LysAsnThrThSerAlaLysLysLysGInGInLysLysLys 420
   |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
520 GGGGGGGAACCAAGGCAAA 541
   |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
420 sGlyLeuIleLeuArgAsnLys 427
   |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:
seq_name: SwissProt_39: COS1_YEAST
seq_documentation_block:
ID COS1_YEAST STANDARD: PRT: 381 AA.
AC P53822:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COS1 PROTEIN.
GN COS1 OR YNL336W OR NO275.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M.;
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DUP/COS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: 271612; CA96270.1; -

DR SGD: S0005280; COS1.

DR InterPro: IPR001142; DUP.

DR Pfam: PF00674; DUP; 2.

KW Transmembrane; Multigene family.

FT TRANSMEM 43 63 POTENTIAL.

FT TRANSMEM 73 93 POTENTIAL.

FT TRANSMEM 232 252 POTENTIAL.

FT TRANSMEM 255 275 POTENTIAL.

SO SEQUENCE 381 AA; 45288 MW; F9AF930CB1935131 CRC64;

alignment_scores:

Quality:	Ratio:	Length:	Gaps:
77.50	1.123	128	6
Percent Similarity: 53.906	Percent Identity: 24.219		

alignment_block:

US-09-823-101-1 x COS1_YEAST ..

Align seg 1/1 to: COS1_YEAST from: 1 to: 381

```

164 ATGACCTTAATATGAAAAAGACAGACATTTCTGTAGTCCATCATGCT 213
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
59 LeuSerValTrpTrpLysLeuSerAsnGlyLysTyTyProLeuIleVa 75
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
214 CAGTGCCACTGAAAAATCCAGCTCTCCACTTTTCTAT..... 250
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
75 IserLeuLeuValLeuPhGInGlyProIlePhGInValLeuValIleCysG 92
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
251 ..ATGCAGAGAAAGACAGAGAGAA.....TATGCCAA 286
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
92 ILeuSerArgLysArgSerLeuSerLysGInLeuIleGInLysLys 108
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
287 AVAGTCACAGATGATTTTG.....TCGAGTGTGATTAAT 321
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
109 GluValThrGluAsnThrProSerSerAspProHisAspTrpGluValVa 125
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
322 GCGTGATCTTGTGCTTTTATTATTTCCAAATTTTATAGACGACAGA 371
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
125 IAlaAlaAsnLeu.....AsnSerTyL 133
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
372 AATACTTTAAGAAATTTCTGACTACAAAAATTTGTTTATATCCCG 421
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
133 euTyTyGInLunLysAlaTrpAsnThrLysTyTyPhGInHisnAlaMet 149
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
422 ..TGCCAA...GTTTCAGAAATATGCAATATGAAACAGATTAATAA 465
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
150 ValCysGInGluAlaPhGInArgThrThrLeuLeuGluProPhoSerLeu 166
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
466 GAAGAAACTACTCAAGTTAAAAAACCAAAA 499
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
166 sLysAspGluAlaAlaLysValLysSerPheLys 177
   :|||: |||: |||: |||: |||: |||: |||: |||: |||:
seq_name: SwissProt_39: COS7_YEAST
seq_documentation_block:
ID COS7_YEAST STANDARD: PRT: 383 AA.
AC Q07788:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COS7 PROTEIN.
GN COS7 OR YDL248W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

```

RP SEQUENCE FROM N.A.
 RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wagner G.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DUP/COS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC EMBL: Z74296; CAA9828.1; -
 DR SGD: S0002407; COS7.
 DR InterPro: IPR001142; DUP.
 DR Pfam: PF00674; DUP; 2.
 DR Transmembrane; Multigene family.
 KW TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 SQ SEQUENCE 383 AA; 45679 MW; 5DC6695D06A61663 CRC64;

alignment_scores:

Quality: 76.50 Length: 128
 Ratio: 1.109 Gaps: 6
 Percent Similarity: 53.906 Percent Identity: 22.656

alignment_block:

US-09-823-101-1 x COS7_YEAST ..

Align seg 1/1 to: COS7_YEAST from: 1 to: 383

```

164 ATGACCTTAATGAGAAACAGACAGACATCTTGTCAGTCCATCATGCT 213
      ::::: ||||| ::::: |||||
59 LeuSerValItrprLysLeuSerAsnScysIleTyrProLeuIleVa 75
214 CAGTGCACGCAAAATCCAGCTCTCCAGTTTCTAT..... 250
      ::||| ::|||
75 lSerLeuValIleuPheLeuGlyProIlePheValIleuValIleCysG 92
251 . ATGCAGAGAAAAAGACAGAGAGAA.....TATGCCAA 286
      ::::: ||||| ::::: |||||
92 lyeuSerAlglysArgSerLeuSerIleGlnPheCysLys 108
287 ATAGTCACAGTATTTTG.....TCTGAGTGGTGAATTAT 321
      ::||| ::|||
109 GluIleThrGlnsIntrProSerSerAspProHisAspTrpGluValVa 125
322 GCGTGATCTTGTTGTTTATTTATTTCAATTTTATAAGACAGCAGA 371
      ::||| ::|||
125 lValAlaAsnLeu.....AsnSerTyrL 133
372 AATACCTTTAAGAAATTCGACACAAAATGGTTTATTTTATTCCTG 421
      ||| ::::: |||||
133 euTyrGluAsnLysAlaTrpAsnThrLysAsnPhePheAsnAlaThr 149
422 ...TGCCAA...GTTCAGAAATATGCCATAAGACAGATACTATAAA 465
      ||::: ::::: |||||
150 AspCysGluLysMetPheArgThrThrValLeuGlnPhePheSerLeu 166
466 GAAGAAAATACCTCAAGGTTAAAAAACCAAA 499
      ||||| ::::: |||||
166 sLysAspLysAlaAlaLysValLysSerPheLys 177

```

seq_name: Swissprot_39:y131_UREPA

seq_documentation_block:

ID Y131_UREPA STANDARD; PRT; 268 AA.
 AC Q9PR13;

DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN U0131.
 GN U0131.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OX NCBL_TaxID-134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3;
 RC MEDLINE=20500219; PubMed-11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum";
 RL Nature 407:757-762(2000).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC EMBL: AF002114; AF30537.1; -
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 SQ SEQUENCE 268 AA; 30686 MW; 01BDC28B5C9A4380 CRC64;

alignment_scores:

Quality: 76.00 Length: 118
 Ratio: 1.246 Gaps: 6
 Percent Similarity: 51.695 Percent Identity: 26.271

alignment_block:

US-09-823-101-1 x Y131_UREPA ..

Align seg 1/1 to: Y131_UREPA from: 1 to: 268

```

80 GAAGGCGTAGACAAATAGAGAAAAAGAGAGAGTCAATAGTG 129
      ||||| ::::: |||||
50 GluIlyPheGlnsAlaArgAlaLeuValAsnSrgInIleIleI 66
130 GGGAAAGGTAGGCCAGAGACAGAGACAGACATGACCTTAATATG. 178
      ||||| ::::: |||||
66 eGlyLysValAlaAspValSerGlyAspAsnHisSerLeuMetSerTrpL 83
179 .....AAAAAGAACAT 190
      |||||
83 euGlyTyrAspGlnHisAlaAsnTyrGlnIleuLeuTyrLysThrIleHis 99
191 TCTGTGTGTA.....GGTCCAGTCATG.....CTGACGGCCATGA 225
      ::::: |||||
100 AspGlnIleuLeuAsnGlyLysThrMetGlnSerLeuIleAsnSnrTras 116
226 AAATCCAGCTCTCCAGTTTCTAT...ATGCAGAGAAAAAGAGAGAA 272
      ::||| ::|||
116 nLysAsnAsnAspProLysLeuTyrGluAlaLeuHisTyrHisAlaHis 133
273 GGAATATGCGCAA.....ATAGTCACAGTATTTTG 304
      ||::: ::|||
133 erLysPheIleAsnHisAsnGlyThrTrpGlyPheValThrLysValLeu 149
305 TCTGAGTGGTGAATTAAGCGTGAATCTTGCTGTTTATATTCGAAT 354
      ||||| ::|||
150 SerAspTrp.....GluAsnSerTrpPheTyrLysTyrAsnGlnLe 163
355 TTTT 358

```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z49661; CAA89694.1; -
 DR SGD: S0003922; COS5.
 DR InterPro: IPR001142; DUP.
 DR Pfam: PF00674; DUP; 2.
 KW Transmembrane; Multigene family.
 FT TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 SQ SEQUENCE 383 AA; 45708 MW; C8E3D0A4AE39462 CRC64;

alignment_scores:
 Quality: 74.50 Length: 128
 Ratio: 1.080 Gaps: 6
 Percent Similarity: 53.906 Percent Identity: 23.438

alignment_block:

US-09-823-101-1 x COS5_YEAST ..

Align seg 1/1 to: COS5_YEAST from: 1 to: 383

```

164 ATGACCTTAATATGTAACAGAACATTCCTGATGTCACATCATGCT 213
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59 LeuSerValIrrPrrPlyrLseuSerAsnAcysIleTyrrProleuIleVa 75
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 CAGTGCACCTGAAATATCATCAGCTCTCCAGTTTCTAT ..... 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
75 lserLeuValrValleuPheleuGlyProIlePheValleuValIleCysg 92
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 . ATGCAGAGAAAAAGACAGAGAA.....TATGCCAAA 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92 lyeuSerIrrGlyrSrrSerLeuSerLysgIleuIleGlnPheCysLys 108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 ATAGTCACGATGTTTG.....TCTGATGCTGATAT 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
109 GluValIrrGlnsInrrProSerSerAspProHisAsprrGluValVa 125
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
322 GGCTGATTCCTGTTCTTTTATATTCAAATTTTATAGACACAGCA 371
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 lAlaIleAsnLeu.....AsnSerIrrL 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
372 AATACCTTTTAAATTCGACTACAAAAATGGTTTATTCCTCG 421
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
133 eutyrrGlnAsnLysAlaTrpAsnThrArgTyrrPhePheAsnAlaMet 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 ...TGGCAA...GTTCAGAAATATGCATATGAGACACAACTACTAAA 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 GlYCysGlnIleuAlaPheArgTyrrThrLeuGlnIleuProPheSerLeu 166
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 GAAGAAAACTACCTCAAGTTTAAAAAACCAAAA 499
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
166 sLysAspGluAlaIleValLysSerPheLys 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: SwissProt_39:PEX2_MOUSE

seq_documentation_block:

ID PEX2_MOUSE STANDARD; PRT; 305 AA.
 AC P55098; 035467;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
 DE PROTEIN 3).
 GN PXPMP3 OR PEX2 OR PAF1 OR PMP35.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SWISS WEBSTER; TISSUE=Liver;

RC MEDLINE=94318312; PubMed=8043297;

RA Wilson G.N., Bryant D.D.;

RT "Structure and expression of mammalian peroxisome assembly factor-1

(PMP35) genes."

RL Biochem. Med. Metab. Biol. 51:140-148(1994).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Cerebellum;

RC MEDLINE=98044231; PubMed=9382874;

RA Faust P.L., Hatten M.E.;

RT "Targeted deletion of the PEX2 peroxisome assembly gene in mice

provides a model for Zellweger syndrome, a human neuronal migration

disorder."

RL J. Cell Biol. 139:1293-1305(1997).

CC - FUNCTION: SOMEWHAT IMPLICATED IN THE BIOGENESIS OF PEROXISOMES.

CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.

CC - SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L27842; AAA21742.1; -

DR EMBL: AF031128; AAB91465.1; -

DR MGD: MGI:107486; Pmp3.

DR InterPro: IPR001841; znf_fing.

DR Pfam: PF00097; zf-C3HC4; 1.

DR SMART: SM00184; RING; 1.

DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.

KW Transmembrane; Peroxisome; Zinc-finger.

FT TRANSMEM 140 159 POTENTIAL.

FT TRANSMEM 195 213 POTENTIAL.

FT ZN_FING 244 283 RING-TYPE.

FT CONFLICT 2 8 AAREEST -> TGKEEM (IN REF. 2).

FT CONFLICT 85 85 H -> Y (IN REF. 2).

FT CONFLICT 91 94 PMPV -> LNL (IN REF. 2).

FT CONFLICT 101 101 N -> T (IN REF. 2).

FT CONFLICT 105 105 L -> W (IN REF. 2).

FT CONFLICT 114 114 R -> K (IN REF. 2).

FT CONFLICT 138 138 C -> F (IN REF. 2).

FT CONFLICT 142 142 V -> L (IN REF. 2).

FT CONFLICT 152 152 M -> I (IN REF. 2).

FT CONFLICT 225 227 TLC -> IPL (IN REF. 2).

FT CONFLICT 295 295 A -> S (IN REF. 2).

FT CONFLICT 298 298 Q -> E (IN REF. 2).

SQ SEQUENCE 305 AA; 34731 MW; FAFB6C8913F959F0 CRC64;

alignment_scores:
 Quality: 74.00 Length: 171
 Ratio: 0.871 Gaps: 8
 Percent Similarity: 49.708 Percent Identity: 21.637

alignment_block:

US-09-823-101-1/rev x PEX2_MOUSE ..

Align seg 1/1 to: PEX2_MOUSE from: 1 to: 305

```

478 GGTAGTTTCTTCTTTAGTATCTGTTCCATTATGCAATATTCGTGA 429
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 GlyGluIleuMetAsnPhelLeuIlePhelGlnLysGlyLysPheAlaTh 165
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 CTTCGCACAG.....GGAATTAATAAACCAATTTTGTAGTCACA 388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
165 lleuThrGlnArgLeuLeuGlnIleHisSer...ValPheCysLysProg 181
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

387 AATCTTAAAGTATTTCTGTGCTGTATATAAATTTGGAAATATATAAA 338
    :: ::::: ::::: ||| ||| :::::
181 llnsmetarglvalglrphcglutyrmetasnargluleu..... 195
337 GAACAAGATGACCATTAATTCACCACTCAGACAAATACCTGTGACTA 288
196 ..... LeuTrpHisGlyPheAlaGluPhe 203
287 TTTTGCATATTTCTGCTGCTGCTTTTCTG..... 256
    : ||| ::::: ||| ::::: :::::
203 e.LeuIlePheLeuProLeuIleAsnIleGlnLysLeuLysAlaLys 219
255 ..... TGCATATAGAAAACGTGAGAGCTGATTTTTCAGTGC 218
    ||| ||| ||| :::::
220 LeuSerSerTrpCysThrLeuCysThrGlyAlaAlaGlnHisAspSerTh 236
217 ACTGAGCATGAGTGCACCTACACAAAGATGTTCTGTTCCATATTAAAG 168
    ::::: ||| ||| :::::
236 rLeuGlySerSergly.....LysGluCys..... 244
167 TCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
    ::::: ||| ::::: ||| ::::: ||| :::::
245 ..... AlaLeuCysGlyGluTrpProThrMetProHisThrIleGly 258
117 CTCTTTTCTCTTTTCTCTATTTTGTCTAGCCCTCCCTCCCTCACTTAC 68
    ::::: ||| ::::: ||| :::::
259 CysGlnHisValrPheCysTrpTrpCysValLysSerSer.....PheLe 273
67 CTTAGAGAGTCTGCTGCTGCTTACCATTAATGATGAGCTTCACTCAGATCTC 18
    ::::: ||| ||| ::::: ||| :::::
273 upheAspIleTrpPheThrCysProLysCysGlyThrGlnValHisSerV 290
17 TGGTTCCAATG 7
    :: ||| ::
290 alGlnProLeu 293

seq_name: swissprot_39:HUNK_HUMAN

seq_documentation_block:
ID HUNK_HUMAN STANDARD; PRT; 714 AA.
AC P57058;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HORMONALLY UPREGULATED NEU TUMOR-ASSOCIATED KINASE (EC 2.7.1.1.-)
DE (SERINE/THREONINE PROTEIN KINASE MAR-V) (BI9).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA SCOTT H.S., Antonarakis S.E., Papasavvas M.P., Michaud J.;
RT "A putative serine/threonine protein kinase MAR-V on human chromosome
  21q22.1".
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE SRR/THR FAMILY OF PROTEIN KINASES.
CC SNF1 SUBFAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
DR EMBL: AJ271722; CAB7146.1; -
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00069; pkinase; 1.

```

```

DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 62 320 PROTEIN KINASE.
FT NP_BIND 68 76 ATP (BY SIMILARITY).
FT BINDING 91 91 ATP (BY SIMILARITY).
FT ACT_SITE 186 186 BY SIMILARITY.
SQ SEQUENCE 714 AA; 79685 MW; 87FC85C67DB38AA7 CRC64;

alignment_scores:
    Quality: 74.00 Length: 121
    Ratio: 1.156 Gaps: 5
Percent Similarity: 52.893 Percent Identity: 23.967

alignment_block:
US-09-823-101-1 x HUNK_HUMAN ..

Align seg 1/1 to: HUNK_HUMAN from: 1 to: 714

203 CCACATCATGCTCAGTGCACCTGAAATCCAGCTTCCAGTTTCTATAT 252
    ||| ::::: ||| ||| ||| ::::: ||
335 ProAsnArgIleSerLeuGlnAspLeuSerProSerValValLeuHisMe 351
253 GCAGAGAAAGAGACAGAGAGAAATATGCCAAATAGACAGATTT 302
    | ::::: || ||| ::::: :::::
351 tThrGlnLys..LeuGlyTrpLysAsnSerAspValIleAsnTrpVal 367
303 TGCTGAGTGTGTAATTAATGCTGATCTTGTCTTTTATATTTCGA 352
    ||| ::::: ||| ::::: ||| :::::
367 euserAsn.....ArgAlaCysHisIleLeuAlaIleTrpPheLeu 380
353 ATTTTATTAAGACGACAGAAATACTTTTAAGATTTCTGACTACAAA 402
    ::::: ||| ::::: ||| :::::
381 LeuAsnLysLysLeuGlnArgTrp.....LeuSerGlyLys 393
403 TTGCTTTTAAATTCCTGCTGC.....CAAGTTGAGA 434
    : ::::: ||| ::::: ||| :::::
393 rAspIleGlnAspSerLeuCysTrpLysThrArgLeuTrpGlnIleGlu 410
435 AATATGCCATTAATGAGACA.....GATAACTTAAAG 466
    ||| ||| ::::: ||| :::::
410 yStyArgAlaProLysGlnSerTrpGlnAlaSerLeuAspThrTrpThr 426
467 AAGAAATACCTCTCAAGTTAAAAAACCACAAAAAATAAAAA 516
    ::::: ||| ||| ::::: ||| :::::
427 ArgAspLeuGlnPheHisAlaValGlnAspLysProLysGlnGln 443
517 AAGGGGGGGGGA 529
443 uLysArgGlyAsp 447

seq_name: swissprot_39:PEX2_RAT

seq_documentation_block:
ID PEX2_RAT STANDARD; PRT; 305 AA.
AC P24392; O63733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
  PROTEIN 3).
GN PAMP3 OR PEX2 OR PAF1 OR PMP35.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=SPRAGUE-DAWLEY;
RC MEDLINE=91163637; PubMed=1750930;

```

DR	EMBL; X57988; CAA41054.1; -	
DR	EMBL; D30616; BAA06306.1; -	
DR	EMBL; D30617; BAA06307.1; -	
DR	PIR; S14290; S14290.	
DR	InterPro; IPR001841; ZnF_RING.	
DR	Pfam; PF00097; zf-C3HC4; 1.	
DR	SMART; SMO0184; RING; 1.	
DR	PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.	
KT	Transmembrane, Peroxisome; Zinc_finger.	
FT	TRANSMEM 140 159	POTENTIAL.
FT	TRANSMEM 195 213	POTENTIAL.
FT	ZN_FING 244 283	RING-TYPE.
FT	VARIANT 125 125	R->P.
SO	SEQUENCE 305 AA; 34767 MW; 249058A55A8E854 CRC64;	

Align seg 1/1 to: PEX2_RAT from: 1 to: 305

```

217 ACAGGATAGTAGGACCTACACAGATCTTCGTTCCATTTAAG 168
    :|||||
236 rleuglyserSergly.....lysGlucyS..... 244
    :|||||
167 TCATGTGCTCTGCTGCTGCTGGCTACCTTTCCCAATGAGCT 118
    :|||||
245 .....AlaLeucySgLyIuTrIprProHmetProHISThrIleGly 258
    :|||||
117 CTTCTTTTCTCTTTTCCATTTTGTGTAGCCCTCCCTCAGCTTAAC 68
    :|||||
259 CysGlnHisIalPheCysTyTyTyCysValIyserSer.....Phele 273
    :|||||
67 CTTAGAGATCTCTGCTGTTGTTACCATAGTGTAGATCTCAGAGATCTC 18
    :|||||
273 upheaPmetTyTrPhehncycProIyScySgLyThGlIvalHISserV 290
    :|||||
17 TGGTTCACATG 7
    :|||||
290 alGlnProIeu 293
    :|||||

seq_name: SwissProt_39:HOFC_HAEIN

seq_documentation_block:
ID HOFC_HAEIN STANDARD: PRT; 406 AA.
AC P44621;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TRANSPORT PROTEIN HOFC HOMOLOG.
GN HOFC OR HOFC OR HI0297.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
    |||
RN SEQUENCE FROM N.A.
RP STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shilly R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fline L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.W., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RT Science 269:496-512(1995).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -1 SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XCPS FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U37315; AAC21961.1; -.
CC DR TIGR: H10297;
CC DR InterPro: IPR001992; Bact_sec_systII.
CC DR Pfam: PF00482; GSPIL_F_2.
CC DR PROSITE: PS00874; T2SP_F_1.
CC KW Transport; Transmembrane; Inner membrane; Complete proteome.
CC FT TRANSMEM 167 187 POTENTIAL.
CC FT TRANSMEM 214 234 POTENTIAL.
CC FT TRANSEM 379 399 POTENTIAL.
CC SQ SEQUENCE 406 AA; 46290 MW; 429D83B7C10F8F82 CRC64;

```


301 AATCACTGTGACTAATTGGCATATTTCCCTGCTGCTTTTCCTGCA 252
354 SerGly 355
251 TATAGAAACTGGAGAGCTGGATTTTCAGTGGCACTGAGCATGAGTGA 202
356 TrpValAlaAspGlnAsnTrpIleLysLysTyrHis..... 367
201 CCTACACAGAGATGTTCTGTTTCCATATTAAGTCATGTCGTCTGCC 152
368 . TyrHisLysSerTyrLeuIle..... LeuCysG 377
151 TGCTCTCTGGCCTACCTTTCCCACTAATGACCTCTCTTTTCTCTTT 102
377 LysIleThrAsnLeu.LeuAlaProLeuAlaThrThr..... Ph 389
101 TCCTATTTTG 92
389 eProLeuLeu 392


```

439 TGGCCTATGTGAAACAAGATACTAAAGAGAAACTCTCAAGCTAA 488
371 ...ValIaspPProGlnaspValPheIysProThrThrSerArgLeuL 386
489 AAAAACCA 496
||:::||
386 ysgIlnPro 388

seq_name: sp_bacteria:Q9JVL6

seq_documentation_block:
ID Q9JVL6 PRELIMINARY; PRT; 168 AA.
AC Q9JVL6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POSSIBLE MEMBRANE PROTEIN.
GN NMA0786
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria
CX NCBI_taxonomy:65699;

```

RP SEQUENCE FROM N.A.
RC STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222555; PubMed-10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
RA Jagsels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84069.1; -.
KW Complete proteome.
SQ SEQUENCE 168 AA; 20297 MW; 91A6E1EE17733412 CRC64;

```

alignment_scores:      length: 174
                       quality: 86.50
                       ratio: 1.095
                       Gaps: 9
Percent Similarity: 45.402  Percent Identity: 25.287

alignment_block:
US-09-823-101-1/rev x Q9JVL6 ..

Align seg 1/1 to: Q9JVL6 from: 1 to: 168

```

[illegible]

```

288 ATTGGGACATATTCCTTCGTGCTCTTTTCTGACATATAGAAACTGG 239
      |||:::| |||:::| |||:::| |||:::|
93 IlePheLysTYrThmPhePro_SerTYrPheLysLeuIleasnLysAsm 109
      |||:::| |||:::| |||:::| |||:::|
238 AGAGCTGGATTTTTCAGTGGGACAGCATGAGCACTACACAAAGAT 189
      |||:::| |||:::| |||:::| |||:::|
109 eLargTYrLeuPhe.....HisMetLsn 116

188 GTT.....CTGTTTCATATTAAGTCATGTGCTCTGCGCTCTC 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 ValPheCysLeuPhePheIleCysPheSerIleTYrLeu.....Ileasn 132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 TGGCTACCTTCCCACTAATGACACTCTTTTTCCTTTTCCAT 95
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 sely.....GluasnGlnTYrIleAlaTYrPheSerCysTYrPheP 146

94 TTGTTCTAGCCCTTCCCTC 75
      || ||| ||| ||| ||| ||| ||| ||| ||| |||
146 heGlyLeuThrTYrProLeu 152

```

	seq_documentation_block:		
ID	089906	PRELIMINARY;	PRT; 3070 AA.
AC	089906;		
DT	01-NOV-1998 (TReMBLrel. 08, Created)		
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)		
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)		
DE	FUSION PROTEIN.		
OS	Beet yellow virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;		
OC	Closterovirus.		
CC	NCBI_Taxid=12161;		
CX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=CALIFORNIAN;		
RC	MEDLINE=98285746; PubMed=9621048;		
RX	Peremyslov V.V., Hagihara Y., Dolta V.V.;		
RA	"Genes required for replication of the 15.5-kilobase RNA genome of a		
RT	plant closterovirus."		
RL	J. Virol. 72:5870-5876(1998).		

RP	SEQUENCE FROM N.A.					
RC	STRAIN-CALIFORMITAN.					
RA	Dolja V., Peremyslov V., Hagivara Y.;					
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AF056575; AAC25115.1;	-				
DR	InterPro: IPR000606; Viral_helicaseI.					
DR	InterPro: IPR001788; RNA_dep_RNAPol2.					
DR	InterPro: IPR002588; V_methyltransf.					
DR	Pfam: PF00978; RNA_dep_RNAPol2; 1.					
DR	Pfam: PF01443; Viral_helicaseI; 1.					
DR	Pfam: PF01660; Vmethyltransf; 1.					
FT	CHAIN	1	588			
FT	CHAIN		2042			
FT	PAPAII-LIKE LEADER PEPTIDE.					
FT	PUTATIVE METHYLTRANSFERASE AND RNASE HELICASE.					
FT	CHAIN	2626	3070	RNA-DEPENDENT RNA POLYMERASE.		
QO	SEQUENCE	3070 AA:	345803 MW;	6FC307093FIDF36E CRC64;		

```
alignment_scores:
  quality: 83.50      length: 147
  ratio: 1.14         gaps: 7
Percent Similarity: 49.660  Percent Identity: 25.170
```

```

alignment_block:
US-09-823-101-1/rev x 089906    ..

Align seg 1/1 to: 089906 from: 1 to: 3070

```

448 CATTATGCGCATATTTCTGCACCTTGCCACAGGAAATTAATAAACCAATTTTT 359
||||| :: |||||
44 HisphaeAgLeuNhrlSeuSerLeuAapHseValProArgProLeuPh 60

```

398 TTGTAGTCAGAAATTCCTTAAAGATATTTCTGTCTTATAAAATTGG 349
| : : :
60 eleusnAlaArgVal..... 65
348 AATATATAAAGACAAAGATCAGCCATATTCACACCTAGACAAAT 299
: : : : :
66 .....HisLeuArgAlaSerThrAspAsnProLeu..... 75
298 CACTGTGACTATTTGGCATATTTCTCTCTCTCTCTCTCTGATAT 249
| : : : :
76 .....ProSerLeuProLeuGlyPheHisAl 84
248 AGAAACTG.....AGACCTGATTTTTCAGTG 220
| : : : :
84 aGluThrPheValLeuGluLeuAsnGlySerSerAlaProPheSerIleP 101
101 roSerArgHisIleAspPheValValAsnArgProPheSerValPhePro 117
175 TATTAAGTCATGTGTCCTCTGCTCTGCTCTGCTCTGCTCTTCCCACT 126
: : : : :
118 .....ThrGluValLeuSerValSerSerLeuArgThrProSerArgIle 132
125 AATGACCTCTCTTTTCTCTTTTCCATTTTGTGTCTAGCCCTCCCT 76
| : : : :
132 upheAlaLeuLeuCysAsp.PhePheLeuTyrCysSerIysProGlyPro 148
75 CACTTAACCTTAGAGATCTGCTGCTGTGTACCATATGT 37
: : : : :
149 CysValGluIleAlaSerPheSerThrProProCys 161

seq_name: sp_virus:Q90703
seq_documentation_block:
ID Q90703 PRELIMINARY; PRT: 3090 AA.
AC Q90703:
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE P348.
OS Beet yellows virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=12161;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BYV-4;
RA MEDLINE=20079557; PubMed=10611288;
RA Perezhyslov V.V., Hagihara Y., Dolja V.V.;
RT "HSP70 homolog functions in cell-to-cell movement of a plant virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BYV-4;
RA Perezhyslov V.V., Hagihara Y., Alphanova D., Dolja V.V.;
RL Submitted (sep-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF140581; AAF14300.1;
DR InterPro; IPR000606; Viral_helicase1.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR002588; V_methyltransf.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
SQ SEQUENCE 3090 AA; 347990 MW; 09C779A57DEB74B9 CRC64;

```

alignment_scores: Quality: 83.50 Length: 147
Ratio: 1.144 Gaps: 7
Percent Similarity: 49.660 Percent Identity: 25.170

```

alignment_block:
US-09-823-101-1/rev x Q90703 ..
Align seg 1/1 to: Q90703 from: 1 to: 3090
448 CATTATGGCATTTTCTGACTTGGCACAGGAATTTAAAAACATTTT 399
| : : : :
44 HisPheArgLeuThrLeuSerLeuAspPheSerValProArgProLeuPh 60
398 TTGTAGTCAGAAATTCCTTAAAGTATTTCTGTGCTTATAAAATTGG 349
| : : : :
60 eleusnAlaArgVal..... 65
348 AATATATAAAGACAAAGATCAGCCATATTCACACCTAGACAAAT 299
: : : : :
66 .....HisLeuArgAlaSerThrAspAsnProLeu..... 75
298 CACTGTGACTATTTGGCATATTTCTCTCTCTCTCTCTGATAT 249
| : : : :
76 .....ProSerLeuProLeuGlyPheHisAl 84
248 AGAAACTG.....AGACCTGATTTTTCAGTG 220
| : : : :
84 aGluThrPheValLeuGluLeuAsnGlySerSerAlaProPheSerIleP 101
101 roSerArgHisIleAspPheValValAsnArgProPheSerValPhePro 117
175 TATTAAGTCATGTGTCCTCTGCTCTGCTCTGCTCTGCTCTTCCCACT 126
: : : : :
118 .....ThrGluValLeuSerValSerSerLeuArgThrProSerArgIle 132
125 AATGACCTCTCTTTTCTCTTTTCCATTTTGTGTCTAGCCCTCCCT 76
| : : : :
132 upheAlaLeuLeuCysAsp.PhePheLeuTyrCysSerIysProGlyPro 148
75 CACTTAACCTTAGAGATCTGCTGCTGTGTACCATATGT 37
: : : : :
149 CysValGluIleAlaSerPheSerThrProProCys 161

seq_name: sp_Invertebrate:Q9N8W5
seq_documentation_block:
ID Q9N8W5 PRELIMINARY; PRT: 99 AA.
AC Q9N8W5:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, last annotation update)
DE PROBABLE SIMILAR TO RING-H2 FINGER PROTEIN RHA1A.
GN CHRL119.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TREU927;
RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,
RA Leonard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,
RA Gerrard C., Rajadream M.A., Barrell B.G.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL359782; CAB95418.1;
SQ SEQUENCE 99 AA; 11941 MW; B66DA5F683CF064 CRC64;

```

alignment_scores: Quality: 82.50 Length: 32
Ratio: 3.438 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 53.125

alignment_block:
US-09-823-101-1/rev x Q9N8W5 ..


```

156 GCGCCGCTCTCTGCGCCACCTTTCCCAATAGACCTCTCTTTTCT 107
164 ..... ValPThrIlePheIle 169
106 CTTTTCCTATTTTGTTCAGCCCTCCCTCCTACCTTAC..... 68
170 ILETyrAlaPheIleAsnTyrProPheGlyIleAsnAsnLysAsnAspVa 186
67 .....CTTAGAGTTTGTGCTTGTTCACCAATATGTCAGTCTCA 28
186 IcySyrThrLeuAlaArgPheSerAsnGlyMetGluLeuValaIlePhe 203
27 CACAGATC 20
203 erThrIeu 205

seq_name: sp_bacteria:Q06370

seq_documentation_block:
ID Q06370 PRELIMINARY; PRT; 203 AA.
AC Q06370:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE REPRESSOR PROTEIN OF SURFACE PROTEIN ANTIGEN GENE.
GN PAR.
OS Streptococcus downei (Streptococcus sobrius).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT3791;
RX MEDLINE=93322312; PubMed=8331066;
RA Takahashi I., Okanashi N., Hamada S.;
RT "Molecular characterization of a negative regulator of Streptococcus
RT sobrius surface protein antigen gene.";
RL J. Bacteriol. 175:4345-4353(1993).
CC -I- FUNCTION: THIS PROTEIN IS A REPRESSOR OF THE SURFACE PROTEIN
CC ANTIGEN GENE.
DR EMBL: D13323; BAA02581.1;
KW Transcription regulation; Repressor; DNA-binding.
SQ SEQUENCE 203 AA; 23395 MW; 9793DB7B40FEFB37 CRC64;

alignment_scores:
Quality: 80.50 Length: 181
Ratio: 1.134 Gaps: 10
Percent Similarity: 39.227 Percent Identity: 25.414

alignment_block:
US-09-823-101-1/rev x Q06370 ..

Align seg 1/1 to: Q06370 from: 1 to: 203

```

```

78 ..... 78
330 GAATGACCATTAATTACACACTGACAGAAAATCAGTGTACTATTTGGC 281
79 ..... HisGlyPheThrPheLys 84
280 ATATTT...CCTTCCTGTCCTTTTCTGCTGATATATAAAAGTCAGAGC 234
85 LeuPheLysProGluAsnProLeuLys...TyrGlyThrAspTrpArgLe 100
233 TGATATTTTCAGTGCAGTGCAGATGATGAGCCTACACAGAAATGTTCT 184
100 uGlyLeuPheAsnAlaThrSerProGlySerArgTyrProArg..... 114
183 GTTTTCATATTAAGTCATGTCGTCTGCGCTGCTGCTGACCTACCTT 134
115 .....SerGlnMetLeuIleIleTyrIeu 122
133 TGCCCA...CTAATGACCTCTCTTTTCTCTTTTCCATTTTGT.... 91
123 AlaProPheValIeuThrSerLeu...LeuLeuThrIleuLeuAlaLeuG 139
90 .....TCTAGCCCTTCCCTCCTCACTTAACCTTAGAGGTTCTG 55
139 IYThrLeuSerProLeuAlaTyrIleuPheLeuAlaValIle 152

seq_name: sp_bacteria:Q9KC34

seq_documentation_block:
ID Q9KC34 PRELIMINARY; PRT; 457 AA.
AC Q9KC34:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE BH1740 PROTEIN.
GN BH1740.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kubara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001513; BAB05459.1;
KW Complete proteome.
SQ SEQUENCE 457 AA; 51577 MW; 054532562FF36386 CRC64;

alignment_scores:
Quality: 80.50 Length: 195
Ratio: 0.885 Gaps: 9
Percent Similarity: 46.667 Percent Identity: 24.615

alignment_block:
US-09-823-101-1/rev x Q9KC34 ..

Align seg 1/1 to: Q9KC34 from: 1 to: 457

```



```

248 AGAAACTGAGAGCTGATTTTTCAGTGCACATGACATGATGAGACCT 199
    ::::::::::::::|
40 SerArgLeuSerSerAlaSerPhe..... 47
198 ACACAGAATGTTCTGTTTCCATATTAAAGTCATGCTGCTGCTCC 149
    |||||
48 .....LeupheSerLeuThrLysSer..... 54
148 TCTCTGGCTACCTTTCCCACTAATGACCTCTCTTTTCTCTTTCC 99
    |||
55 .....GlnProGlnThrGlnSerPhePheValPheSer 66
98 TATTTGTTCTAGCCCTTCCC 78
    |||||
67 LeupheLeuArgAlaAsnPro 73

```

seq_name: sp.vertebrate:073658

```

seq_documentation_block:
ID 073658 PRELIMINARY; PRT; 557 AA.
AC 073658:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OLGRK-C.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198545; PubMed=9539142;
RA Hisatomi O., Matsuda S., Satoh T., Kotaka S., Imanishi Y.,
RA Tokunaga F.;
RA "A novel subtype of G-protein-coupled receptor kinase, GRK7, in
RT teleost cone photoreceptors.";
RL FEBS Lett. 424:159-164(1998).
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES. GPRK
CC SUBFAMILY.
DR EMBL; AB009568; BAA25670.1; -.
DR HSSP; P05132; ICRP.
DR InterPro: IPR000719; Euk.pkinase.
DR InterPro: IPR000239; GPCR_kinase.
DR InterPro: IPR000961; Pkinase_C.
DR InterPro: IPR000342; RGS.
DR pfam; PF00069; pkinase.1.
DR PRINTS; PR00717; GPCR_KINASE.
DR SMART; SM00315; RGS.1.
DR SMART; SM00220; S_TKC.1.
DR SMART; SM00133; S_TK_X.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM.1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST.1.
DR PROSITE; PS00132; RGS.1.
DR ATP-binding; Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 557 AA; 63656 MW; A968753332E0065A CRC64;

```

alignment_scores:

Quality:	79.50	Length:	74
Ratio:	2.038	Gaps:	3
Percent Similarity:	52.703	Percent Identity:	33.784

alignment_block:

US-09-823-101-1 x 073658 ..

Align seg 1/1 to: 073658 from: 1 to: 557

```

344 TATTCGAATTTTATAGA.....GC 366
    |||||
166 TyrPheAspLysPheLeuGlnTrpLysGlnTrpLysGlnProIle 182
    ::::::::::::::|
367 ACACGAATCTTTTAAAGATTTCTGACTACACAAAATGCTTTTAA 416
    ::::::::::::::|
182 rAspLysTyrPheTyrGlnPheArgThrLeuGlnLysGlnLysPheGly 199
    ::::::::::::::|
417 CCCGTGGCAGCTCAGAAATATGCCATTAATGAAACAAGATTAAC 466
    |||||
199 luvAlCysAlaValGln..... 204
467 AGAAACTACCTCAGGTTA.....AAAAACCAAAAAAAAAA 507
    |||||
205 ValLysAsnThrGlnGlnMetTyrAlaCysLysLysLeuCysLys 221
    ::::::::::::::|
508 AAAAAAAAAAAGGGGGGCA 529
    ::::::::::::::|
221 glEuLysLysLysGlyGlyGlu 228

```

seq_name: sp.fungi:004119

```

seq_documentation_block:
ID 004119 PRELIMINARY; PRT; 674 AA.
AC 004119:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE D9461.3/P.
GN PHM5 OR D9461.37 OR YDR452M.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX Dietrich F.S.;
RA Dietrich F.S.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.;
RA Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Jia Y., Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; U33007; AAB64872.1; -.
DR SGD; S0002860; PHM5.
KW SEQUENCE 674 AA; 78344 MW; EEC78BC7568098B3 CRC64;

```

alignment_scores:

Quality:	79.50	Length:	110
Ratio: <td>1.242 <td>Gaps: <td>6 </td></td></td>	1.242 <td>Gaps: <td>6 </td></td>	Gaps: <td>6 </td>	6
Percent Similarity: <td>58.182 <td>Percent Identity: <td>29.091</td> </td></td>	58.182 <td>Percent Identity: <td>29.091</td> </td>	Percent Identity: <td>29.091</td>	29.091

alignment_block:

US-09-823-101-1 x 004119 ..

Align seg 1/1 to: 004119 from: 1 to: 674

```

256 GAGAAAAGCAGAGAGAA.....ATATGCCAAATAGTCACAG 296
    |||||
463 GluLysAspLysLysLysLysProLleThrArgLysGluLeu 479
    ::::::::::::::|
297 TGATTTTCTGAGTGGAATTATGCTGATTCCTTTTATAT 346
    ::::::::::::::|
479 egluArgTyrSerIleValAsnIleGlyGlySerValIleProThrPhea 496

```

```

347 TTCGAATTTT.....TATAGACGACGAAATCTT..... 378
|||||
496 snProSerPheArgIleTrpGluTyrAsnIleThrAspIleValAsnsp 512
379 .....TATAGAAATTTGCTACATCAAAAAATTG.....TTTTTAAT 416
|||||
513 SerAsnIleGluValSerGluTrpAspSerLeuGluAspGluMetGluSps 529
417 CCCTGTGCAAGTTCAGAAATATGCCATATGCAACAAGAT..... 457
|||||
529 erLeuAsnIleGluValSerGluAspSerLeuGluAspGluMetGluSps 545
458 ...AACTAAAGAGAGAAAACTACTCAAGCTTAAAAAACCAAAAAAAA 504
|||||
546 SerAsnIleGluValGlyLeuAsnArgGluTrpMetGlyGluTrpSps 562
505 AAAAAAAAAAAAAAAAAAGGGGGGGAACA 532
|||||
562 nIlySlySlySlySlySlySlySlySlySlySlySlySlySlySlySly 571

```

seq_name: sp_Invertebrate:Q20109

seq_documentation_block:

```

ID Q20109 PRELIMINARY; PRT; 1398 AA.
AC Q20109;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F36H1.2 PROTEIN.
CN F36H1.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68760; CA92996.1; -.
DR HSSP; Q00421; IAWC.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00023; ank.10.
DR PRINTS; PRO1415; ANKRTIN.
DR SMART; SM00248; ANK; 9.
DR PROSITE; PS00178; AA_tRNA_LIGASE_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 7.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 1398 AA: 156568 MW; D5080A9E89AB1BEC CRC64;

```

alignment_scores:

```

Quality: 79.00 Length: 63
Ratio: 2.026 Gaps: 2
Percent Similarity: 61.905 Percent Identity: 33.333

```

alignment_block:

US-09-823-101-1/rev x Q20109 ..

Align seg 1/1 to: Q20109 from: 1 to: 1398

```

519 TTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTTTGTACCTTGAGGTGTTT 470
|||||
491 PhePheAlaIlePheLeuPhePheGlyMetPheSerLeuTrpPheThrMe 507
469 CTTCCTTTAGTTATCTGTTCATATATGCAATATTCGAACCTTGACACA 420
|||||
507 tleuIleAlaIleSerAsnSerValThrAlaTrpIleuIleSerTrpSps 524
419 GGGAAATTAATAAACCAATTTTGTAGTCAGAAATTCCTAAAGTATTTTC 370
|||||
524 alPheLeu.....LeuIlePheIleIle.....Phe 532
369 TGTGCTTTATATAAATTTGGAAATATATAAAGAACAA 331
|||||
533 CysSerLeuIleValValValTrpTrpGlyAspArgLys 545

```

```
US-09-823-101-2 x AAE02399 ..
Align seg 1/1 to: AAE02399 from: 1 to: 382

39 CAAAGAGTAGAGAGATAGAGTCAGGCTCTATACCAATACTACTGCG 88
   ::::::::::::::::::::
311 ArgGlyIylArgIylArgIylArgIyl..... 317
89 CTTAACTCCCGAGGAAGGAGATCCAAACCTGGAAGACAGACGA 138
   ::::::::::::::::::::
318 ...ArgSerArgGlyArgGlyIyl.....GlyArgArgGlyIylGlys 330
139 GTAAAGTGGG...CGAGTGGGAGGCGAGAAAGCTTGCGTACAGATCAG 185
   ::::::::::::::::::::
330 eArgGAlagIylArgArgGlyGlyGlyArgArgIyl.....Arg 342
186 GGAAGG 191
   ::::::
343 GlyGly 344

seq_name: /SID8/gcdata/geneseq/geneseq/AA2000.DAT:AA51860

seq_documentation_block:
ID AAB51860 standard; Protein: 124 AA.
XX
AC AAB51860;
XX
XX 16-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 34 SEQ ID NO:93.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neutrophic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging.
XX
OS Homo sapiens.
XX
PN WO200061626-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09066.
XX
PR 09-APR-1999; 99US-0128698.
XX
PR 20-JAN-2000; 2000US-0176926.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
PI .Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-619227/59.
DR N-PSDB: AAC93512.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing or ameliorating medical conditions and used for
XX food additives or preservatives -
XX
XX Claim 11; Page 476; 516pp; English.
XX
XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
XX human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
XX AAB51927 represent alternative polypeptides encoded by the genes, and
XX amino acid sequences with which they share homology. The genes, and
XX proteins have activities dependent on the tissues and cells in which they
XX are expressed. Examples of their activities include immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotropic; cerebroprotective; neutrophic; neuroprotective; antibacterial;
XX virucide; fungicide; ophthalmological; and vulnery. The secreted
```

```
CC proteins, polynucleotides, antagonists and agonists may be useful in
CC treating, preventing and/or diagnosing diseases and disorders such as
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
CC used in the isolation and characterisation of the proteins and
CC polynucleotides of the invention.
XX
SQ Sequence 124 AA;
XX
XX
alignment_scores:
Quality: 74.00 Length: 35
Ratio: 3.083 Gaps: 1
Percent Similarity: 68.571 Percent Identity: 42.857

alignment_block:
US-09-823-101-2/rev x AAB51860 ..
Align seg 1/1 to: AAB51860 from: 1 to: 124

185 CTGATCTGTCACCCACCTTCTCGCTCCGCGCCGACCTTCTACGCG 136
   ||| |||||::: ||::||| |||||:::
28 LencysHisProAlaProProSerSerSerProThrSerSer 44
135 CTTGCTCTTCTCC.....AGGTTTGATCTCCCTCTCTCGG 98
   ||| ::||| :: ||:::|::|
44 rLeucysAlaserValSerCysArgIylsTrpValGluProGluArg 61
97 GACTG 93
   |||
61 rGlu 62

seq_name: /SID8/gcdata/geneseq/geneseq/AA2000.DAT:AA43199

seq_documentation_block:
ID AAB43199 standard; Protein: 526 AA.
XX
AC AAB43199;
XX
XX 08-FEB-2001 (first entry)
XX
DE Human ORF2963 polypeptide sequence SEQ ID NO:5926.
XX
XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
XX vulnery; antiporiatic; antiparkinsonian; neutrophic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antinaeemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
```

```
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC77408.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 5111-5112; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertensive; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX SQ Sequence 526 AA;
XX
XX alignment_scores:
XX Quality: 73.00 Length: 51
XX Ratio: 2.920 Gaps: 3
XX Percent Similarity: 49.020 Percent Identity: 43.137
XX
XX alignment_block:
XX US-09-823-101-2/rev x AAB43199 ..
XX
XX Align seg 1/1 to: AAB43199 from: 1 to: 526
XX
XX 190 CCCTGCTGATGCTGACCACTTCTGCGCTCC...ACTGCGCCACT 144
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 27 ProProlLeuSerValProGlnProPheLeuProValPheThrProLe 43
XX 143 TCTACTGCTTGTCTTCTTCACAGGTTTGAGTCCCTTCCCGGAGACT 94
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 43 uLeuSerPro.....SerProAlaProProProI 53
XX
XX 93 GTAAGCAGATGATTTGGCTATGAGAGCCTGACTCTTATCTCTCACT 44
XX ::||| ||||| ||||| |||||
XX 53 leSerPro.....ValLeuProLeuValPro 61
XX
XX 43 CCF 41
XX |||
XX 62 Pro 62
```

```
seq.name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT: AAB42882
seq_documentation_block:
ID AAB42882 standard; Protein: 142 AA.
XX
XX AAB42882;
AC
XX
XX 08-FEB-2001 (first entry)
DT
XX
XX Human ORFX ORF2646 polypeptide sequence SEQ ID NO:5292.
DE
XX
XX Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;
XX vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX W0200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX PA Shinkets RA, Leach M;
XX
XX PI Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX N-PSDB; AAC77091.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4461-4462; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antihypertensive; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
```

XX SQ Sequence 142 AA:

alignment_scores:

Quality:	71.00	Length:	47
Ratio:	2.448	Gaps:	1
Percent Similarity:	61.702	Percent Identity:	36.170

alignment_block:

US-09-823-101-2 x AAB42882 ..

Align seg 1/1 to: AAB42882 from: 1 to: 142

```

3  CTCACACACTGCTGACTGTGAGTGGGGTGCACAGAGGAGGAGGAG 52
   ||| |||:::||||| |||:::|||||:::|||||:::|||||
46  LeuProPromeValAlaSerValAlaGlyLeuGlnAlaGlyLeuAs 62
   ::|||:::|||||:::|||||:::|||||:::|||||
53  AGATAGAGTCAAGGCTCTCATAGCCAAATACTATAGCTTACAGTCCGAG 102
   ::|||:::|||||:::|||||:::|||||:::|||||
62  pGlyGlySerArgGly.....TrpSerGlyG 71

103 GAAGGAGATCCAAACCTTGAGAGACAGACAGCAGTAGA 143
   |||||:::|||||:::|||||:::|||||:::|||||
71  IyArGlyGlnProHisProGlyIyAlaArGlyGlnArg 84

```

seq_name: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:AAE02422

seq_documentation_block:

ID AAE02422 standard; Protein; 661 AA.

XX AC AAE02422:

XX DT 10-AUG-2001 (first entry)

XX DE Fruit fly vasa protein.

XX Vasa; therapy; cancer; oral; brain; ovarian; biliary tract; lung; pancreas; prostate; colorectal; cervical; breast; colon; testis; renal; thyroid; oesophageal; endometrial; gastric; skin; liver; lymphoma; melanoma; glioblastoma; neuroblastoma; neoplasm; medulloblastoma; choriocarcinoma; squamous cell carcinoma; leukaemia; acute lymphocytic; myelogenous; multiple myeloma; Paget's disease; osteosarcoma; Acquired immune deficiency syndrome; AIDS; Bowen's disease; leiomyosarcoma; rhabdomyosarcoma; liposarcoma; Kaposi's sarcoma; fibrosarcoma; tumour; testicular; seminoma; ovarian; dysgerminoma; teratoma; mediastinal; intracranial; fruit fly.

XX OS Drosophila melanogaster.

XX PN WO200136445-A1.

XX PD 25-MAY-2001.

XX PF 16-NOV-2000; 2000WO-US31485.

XX PR 18-NOV-1999; 99US-0166394.

XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

XX PI Castillion DH;

XX DR WPI: 2001-355606/37.

XX PT Novel vasa polynucleotides useful in the diagnosis or treatment of conditions characterized by aberrant expression and/or presence of mutant forms of vasa polynucleotides or polypeptides -

XX PS Claim 4: Page 61-62; 66pp; English.

XX CC The present sequence is fruit fly vasa protein. Vasa is useful in the diagnosis or treatment of conditions characterised by its aberrant expression and/or the presence of its mutant forms. The conditions

CC Include cancers such as biliary tract, brain, breast, colon, ovarian, pancreas, prostate, colorectal, oral, liver, lung, skin, basocellular, testis, renal, thyroid, cervical, endometrial, oesophageal and gastric, lymphomas, melanomas, glioblastomas, neuroblastomas, medulloblastomas, choriocarcinoma, squamous cell carcinoma, haematological neoplasms, acute lymphocytic and myelogenous leukaemia, multiple myeloma, CC Acquired immune deficiency syndrome (AIDS) associated leukaemias, CC Intraepithelial neoplasms, Bowen's disease, Paget's disease, sarcomas CC such as leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, Kaposi's sarcoma and osteosarcoma, tumours such as testicular tumour (eg. seminoma), ovarian tumour (eg. dysgerminoma or teratoma) and tumour of an extragonadal tissue (eg. mediastinal or an intracranial tumour).

XX SQ Sequence 661 AA:

alignment_scores:

Quality:	70.50	Length:	68
Ratio:	2.203 <td>Gaps:</td> <td>3 </td>	Gaps:	3
Percent Similarity:	47.059 <td>Percent Identity:</td> <td>33.824</td>	Percent Identity:	33.824

alignment_block:

US-09-823-101-2 x AAE02422 ..

Align seg 1/1 to: AAE02422 from: 1 to: 661

```

30  GGGGCTGCCAGAGGATAGAGGATAGAGTACAGCTCTCATAGCCAA 79
   ||||| |||||:::|||||:::|||||:::|||||
48  GLyGlyTyrgInGlyGlyAsnArGAsp..... 56
   ::|||:::|||||:::|||||:::|||||:::|||||
80  ATACTATAGCTTACAGTCCCGAGAGGAGGAGATCCA..... 116
   ::|||:::|||||:::|||||:::|||||:::|||||
57  .ValPheGlyArgLLeGlyGlyArgGlyGlyAlaGlyTyra 73
   ::|||:::|||||:::|||||:::|||||:::|||||
117  .....AACCTGAGAGACAAAGCC 137
   ::|||:::|||||:::|||||:::|||||:::|||||
73  rGlyGlyAsnArGAspGlyGlyPheHisGlyGlyArgArgGlyGln 89
   ::|||:::|||||:::|||||:::|||||:::|||||
138  AGTGAAGTGGGAGAGTGGAGGAGCAAGAGTGGGTGACATCAGGG 167
   ::|||:::|||||:::|||||:::|||||:::|||||
90  GlnArgAspPheArg...GlyGlyGlnGlyPheArgGlyGlyGln 105
   |||||
188  AGGG 191
   |||||
105  YGly 106

```

seq_name: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:AAV29514

seq_documentation_block:

ID AAV29514 standard; Protein; 595 AA.

XX AC AAV29514:

XX DT 13-OCT-1999 (first entry)

XX DE Human lung tumour protein SAL-50 1st. predicted amino acid sequence.

XX KW Human; lung tumour protein; therapy; diagnosis; lung cancer; vaccine; immunotherapy; detection; inhibition.

XX OS Homo sapiens.

XX PN WO9938973-A2.

XX PD 05-AUG-1999.

XX PF 26-JAN-1999; 99WO-US01642.

XX PR 22-DEC-1998; 98US-0219245.

PR 28-JAN-1998; 98US-0015022.

PR 28-JAN-1998; 98US-0015029.

PR 18-MAR-1998; 98US-0040828.

PR 18-MAR-1998; 98US-0040831.

PR 23-JUL-1998; 98US-0122191.
PR 23-JUL-1998; 98US-0122192.
XX
XX (CORI-) CORIXA CORP.
PI Frudakis TN, Lodes MJ, Mohamath R, Reed SG;
XX WPI; 1999-479187/40.
DR N-PSDB; AA207194.
XX
XX Lung tumour specific polynucleotides for inhibiting the development
PT of lung cancer
PS
XX Claim 3; Page 151-153; 171pp; English.
XX
XX The present invention describes lung tumour specific polynucleotides
CC and tumour antigens. AA207144 to AA207246 and AA208301 to AA208325
CC represent specifically claimed polynucleotides, and AA29486 to AA29571
CC represent amino acid sequences from the present invention. The lung
CC tumour specific polynucleotides and polypeptides can be used in
CC pharmaceutical compositions and vaccines to inhibit the development of
CC lung cancer. They can also be used to detect lung cancer in a patient.
CC Probes and antibodies derived from the lung tumour sequences are useful
CC in detection of lung cancer.
XX
SQ Sequence 595 AA;

alignment_scores: Quality: 69.00 Length: 47
 Ratio: 2.556 Gaps: 1
Percent Similarity: 57.447 Percent Identity: 42.553

alignment_block:
US-09-823-101-2 x AAY29514 ..

Align seg 1/1 to: AAY29514 from: 1 to: 595

```
30 GGGGCTGTCAGAGAGTAGAGATAGAGTCAGCTCTCATGCCAA 79
||||| :|||:|||||:||||| ||| :|||:
326 GYGlyTTPrglyGlyArgGlyGlySerAlaGlyAlaGlyGly 342
80 ATACTATGGCTTACAGTCCCGAGAGAGGGAGATCCAAACCTGGAAGAA 129
||| :|||:|||||:||||| ||| :|||:
342 yGlyArgGlyGlyArgGlyArgGlyArgGly.....GlyGlyA 355
130 GACAAGCAGTAGAGTGGCGAGTGGAGCGACGAAGGT 170
||| :|||:|||||:||||| ||| :|||:
355 rgGlyGlyGlyGlyAlaGlyArgGlyGlyAlaAlaGly 368
```

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:AA24463

seq_documentation_block:

ID AAB4463 standard; Protein: 595 AA.

XX AAB4463;

XX 05-FEB-2001 (first entry)

XX Human lung tumour-specific antigen encoded by cDNA #100.

XX Lung tumour protein; lung cancer; cytostatic; vaccine.

XX Homo sapiens.

XX WO200060077-A2.

XX 12-OCT-2000.

XX 30-MAR-2000; 2000WO-US08560.

XX 02-APR-1999; 99US-0285323.

PR 09-AUG-1999; 99US-0370838.

PR 30-DEC-1999; 99US-0476235.
PR 03-MAR-2000; 2000US-0518809.
XX
XX (CORI-) CORIXA CORP.
PI Reed SG, Lodes MJ, Mohamath R, Secrist H;
XX WPI; 2000-638466/61.
DR N-PSDB; AAC79147.
XX
XX Novel lung tumor polypeptides and polynucleotides, useful for
PT detecting, monitoring or treating cancer, especially lung cancer -
PS
XX Example 4; Page 167-168; 243pp; English.
XX
XX The present sequence is given in a specification relating to compounds
CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC least an immunogenic part of a lung tumour protein are disclosed.
CC The polypeptides are useful for inhibiting the development of cancer,
CC especially lung cancer. Samples of T cells expressing the polypeptides
CC may be used to inhibit the development of cancer. The polypeptides are
CC also useful for detecting and monitoring the progression of cancer,
CC especially lung cancer.
XX
SQ Sequence 595 AA;

alignment_scores: Quality: 69.00 Length: 47
 Ratio: 2.556 Gaps: 1
Percent Similarity: 57.447 Percent Identity: 42.553

alignment_block:
US-09-823-101-2 x AAB44463 ..

Align seg 1/1 to: AAB44463 from: 1 to: 595

```
30 GGGGCTGTCAGAGAGTAGAGATAGAGTCAGCTCTCATGCCAA 79
||||| :|||:|||||:||||| ||| :|||:
326 GYGlyTTPrglyGlyArgGlyGlySerAlaGlyAlaGlyGly 342
80 ATACTATGGCTTACAGTCCCGAGAGAGGGAGATCCAAACCTGGAAGAA 129
||| :|||:|||||:||||| ||| :|||:
342 yGlyArgGlyGlyArgGlyArgGlyArgGly.....GlyGlyA 355
130 GACAAGCAGTAGAGTGGCGAGTGGAGCGACGAAGGT 170
||| :|||:|||||:||||| ||| :|||:
355 rgGlyGlyGlyGlyAlaGlyArgGlyGlyAlaAlaGly 368
```

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:AA67470

seq_documentation_block:

ID AAY67470 standard; Protein: 250 AA.

XX AAY67470;

XX 12-MAY-2000 (first entry)

XX Np70 protein carboxy terminal region.

XX SH3; np70; src homology 3; nuclear localisation; anticancer;

XX anti-proliferative; anti-inflammatory.

XX Homo sapiens.

XX GB2341182-A.

XX 08-MAR-2000.

XX 01-SEP-1998; 98GB-0019038.

PR 01-SEP-1998; 98GB-0019038.

PA (YAMA) YAMANOUCHI UK LTD.
XX
PI Kellie S. Finan P.
XX
DR WPI; 2000-163495/15.
XX
PT Src homology 3 protein or fragment for preventing or treating
PT proliferative disease, such as cancer or chronic inflammatory disease,
PT comprises src homology 3 domain binding activity and nuclear
PT localization activity -
XX
PS Claim 5; Page 35; 52pp; English.
XX
CC The invention provides a novel protein (np70) that has src homology 3
CC (SH3) domain binding activity and nuclear localisation activity. The
CC proline-rich SH3 binding protein np70 or its fragments are useful for
CC the prevention or treatment of proliferative disease, such as cancer and
CC chronic inflammatory disease. The present sequence represents the
CC np70 protein carboxy terminal region.
XX
SQ Sequence 250 AA;

alignment_scores:
Quality: 68.50 Length: 64
Ratio: 2.635 Gaps: 2
Percent Similarity: 40.625 Percent Identity: 32.812

alignment_block:
US-09-823-101-2/rev x AAY67470 ..

Align seg 1/1 to: AAY67470 from: 1 to: 250

190 CTTCCCTGATCTGTCAACCACTTTCTGCTCCCACT..... 152
|||||
25 ProProLaProProLeuArgProProGlyProProthrhGlyLeuProPr 41
151CGCCCACTTCTACTGCTGTGCTTCTTCAG 121
|||||
41 oglyProProProGlyAlaProProPhleuArgProProGlyMetProG 58
120 GGTTCGATCTCCCTCCCTCCGAGACTGTAAGCATAGTATTGGCTATG 71
|||||
58 lyeuArgGlyProLeuProArg..... 65
70 AGAGCCTGACTTATCTCTACTCTGACACCCCA 29
|||||
66LeuLeuProProGlyProProPro 73

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT: AAY82327

seq_documentation_block:

ID AAY82327 standard; Protein; 641 AA.

XX AAY82327;

XX 21-JUN-2000 (first entry)

DE Human Npw38BP1 transcription related protein SEQ ID NO:1.

KW Human; Npw38BP1; Npw38; transcription related protein; diagnosis;
KW genetic disease; gene therapy; antitumour.

XX Homo sapiens.

XX JP2000060562-A.

XX 29-FEB-2000.

XX 21-AUG-1998; 98JP-0235901.

XX 21-AUG-1998; 98JP-0235901.

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (KATO/) KATO S.
XX
DR WPI: 2000-249677/22.
XX
DR N-PSDB; AAA08075.
XX
PT Human transcription-related protein and a cDNA encoding said protein
PT useful for screening for new antitumor agents and for diagnosis
XX
PS Claim 1; Page 6-7; 10pp; Japanese.
XX
CC The present sequence represents a human transcription-related protein.
CC The present and an antibody against the protein, are useful for
CC diagnosis and treatment of various diseases. The human cDNA is useful
CC as a probe for genetic diagnosis and a gene source for gene therapy.
CC A new antitumor agent can be detected by screening a low molecular
CC compound combining with the protein.
XX
SQ Sequence 641 AA;

alignment_scores:
Quality: 68.50 Length: 64
Ratio: 2.635 Gaps: 2
Percent Similarity: 40.625 Percent Identity: 32.812

alignment_block:
US-09-823-101-2/rev x AAY82327 ..

Align seg 1/1 to: AAY82327 from: 1 to: 641

190 CTTCCCTGATCTGTCAACCACTTTCTGCTCCCACT..... 152
|||||
416 ProProLaProProLeuArgProProGlyProProthrhGlyLeuProPr 432
151CGCCCACTTCTACTGCTGTGCTTCTTCAG 121
|||||
432 oglyProProProGlyAlaProProPhleuArgProProGlyMetProG 449
120 GGTTCGATCTCCCTCCCTCCGAGACTGTAAGCATAGTATTGGCTATG 71
|||||
449 lyeuArgGlyProLeuProArg..... 456
70 AGAGCCTGACTTATCTCTACTCTGACACCCCA 29
|||||
457LeuLeuProProGlyProProPro 464

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT: AAY67469

seq_documentation_block:

ID AAY67469 standard; protein; 641 AA.

XX AAY67469;

XX 12-MAY-2000 (first entry)

DE Np70 protein sequence.

KW SH3; np70; src homology 3; nuclear localisation; anticancer;
KW anti-proliferative; anti-inflammatory.

XX Homo sapiens.

XX GB2341182-A.

XX 08-MAR-2000.

XX 01-SEP-1998; 98GB-0019038.

XX 01-SEP-1998; 98GB-0019038.

XX (YAMA) YAMANOUCHI UK LTD.

PI Kellie S, Finan P;
 XX
 DR WPI: 2000-163495/15.
 XX
 PT SRC homology 3 protein or fragment for preventing or treating
 PT proliferative disease, such as cancer or chronic inflammatory disease,
 PT comprises src homology 3 domain binding activity and nuclear
 PT localization activity -
 XX
 PS Claim 5; Page 35; 52pp; English.
 XX
 CC The invention provides a novel protein (np70) that has src homology 3
 CC (SH3) domain binding activity and nuclear localisation activity. The
 CC proline-rich SH3 binding protein np70 or its fragments are useful for
 CC the prevention or treatment of proliferative disease, such as cancer and
 CC chronic inflammatory disease. The present sequence represents the
 CC np70 protein.
 CC
 XX
 SQ Sequence 641 AA;
 alignment_scores:
 Quality: 68.50 Length: 64
 Ratio: 2.635 Gaps: 2
 Percent Similarity: 40.625 Percent Identity: 32.812
 Alignment_block:
 US-09-823-101-2/rev x AAY67469 ..
 Align seg 1/1 to: AAY67469 from: 1 to: 641
 190 CCTCCCTGATCTGTACCCACCTTCTGCTCCACT..... 152
 ||||| :||| |||||
 416 ProProAlaProProLeuAArgProProGlyProProThGlyLeuProPr 432
 151CGCCACTTCTACTGCTTGTCTTCCAG 121
 |||:|||||
 432 oGlyProProProGlyAlaProProPhLeuAArgProProGlyMetProG 449
 120 GGTTCGATCTCCCTCTGCGGACTGTAAACCATAGATTGGCTATG 71
 |||: :|||
 449 lYLeuAArgGlyProLeuProAArg..... 456
 70 AGAGCCTGACTCTTATCTCTCTACCTCTTGACACCCCA 29
 |||||
 457LeuLeuProProGlyProProPro 464
 seq_name: /SIDB8/gcgdata/geneseq/geneseq/AA2001.DAT:AAV72165
 seq_documentation_block:
 ID AAV72165 standard; Protein: 641 AA.
 XX
 AC AAV72165;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human RNA metabolism protein (RMEP-5).
 XX
 KW Human: RNA metabolism protein; RMEP; nootropic; neuroleptic; antiulcer;
 KW transquilliser; antianemic; antidiabetic; immunosuppressive; cytostatic;
 KW antiaesthetic; antiinflammatory; anti-HIV; human immunodeficiency virus;
 KW antiarthritic; antiatherosclerotic; antiatherosclerotic; antiallergic;
 KW antipneumotoid; antiparkinsonian; antithyroid; nephrotrophic; antiout;
 KW thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;
 KW nervous system disorder; Alzheimer's disease; therapy; gene therapy;
 KW neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;
 KW peripheral nervous system disorder; cirrhosis; cranial nerve disorder;
 KW progressive neural autonomic nervous system disorder; Addison's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; drug screening.
 XX
 OS Homo sapiens.
 XX
 PN WO200078952-A2.

XX
 PD 28-DEC-2000.
 XX
 PF 15-JUN-2000; 2000WO-US16644.
 XX
 PR 17-JUN-1999; 99US-0139922.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PA (BAUG/) BAUGHN M R.
 XX
 PI Bandman O, Yue H, Lal P, Tang YT, Reddy R, Azimzai Y;
 XX
 DR WPI: 2001-102723/11.
 DR N-PSDB: AAD02348.
 XX
 PT New human RNA metabolism proteins (RMEP), useful for diagnosing,
 PT treating, preventing nervous system, cell proliferative,
 PT autoimmune/inflammatory disorders associated with abnormal expression
 PT of RMEP
 XX
 PS Example 1; Page 86-87; 103pp; English.
 XX
 CC The present sequence is human RNA metabolism protein (RMEP-5) encoded by
 CC a cDNA (clone ID 1858421) obtained from PROSNOT18 cDNA library.
 CC Agonists and antagonists of RMEP cDNA are useful for treating diseases or
 CC conditions associated with altered expression of functional RMEP. RMEP
 CC sequence or their mammalian homologues are useful for creating 'knock
 CC out' or 'knock in' humanised animals or transgenic animals to model
 CC human disease. RMEP sequence is useful in the diagnosis, prevention and
 CC treatment of nervous system disorders e.g. Alzheimer's disease, Pick's
 CC disease, Huntington's disease, Parkinson's disease, amyotrophic lateral
 CC sclerosis, and other motor neuron disorders, progressive neural autonomic
 CC nervous system disorders, cranial nerve disorders, spinal cord diseases,
 CC muscular dystrophy and other neuromuscular disorders, peripheral nervous
 CC system disorders, mental disorders including anxiety and schizophrenia,
 CC anorexia etc, cell proliferative disorders e.g. actinic keratosis,
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), cancers e.g. adenocarcinoma, leukemia, lymphoma,
 CC melanoma etc., and autoimmune/inflammatory disorders such as acquired
 CC immuno deficiency syndrome (AIDS), Addison's disease, allergies, anaemia,
 CC asthma, diabetes mellitus, rheumatoid arthritis, Grave's disease and
 CC autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APCED).
 CC RMEP cDNA is useful for somatic or germline gene therapy. RMEP sequence
 CC is useful several drug screening assays.
 CC
 XX
 SQ Sequence 641 AA;
 alignment_scores:
 Quality: 68.50 Length: 64
 Ratio: 2.635 Gaps: 2
 Percent Similarity: 40.625 Percent Identity: 32.812
 Alignment_block:
 US-09-823-101-2/rev x AAY72165 ..
 Align seg 1/1 to: AAY72165 from: 1 to: 641
 190 CCTCCCTGATCTGTACCCACCTTCTGCTCCACT..... 152
 ||||| :||| |||||
 416 ProProAlaProProLeuAArgProProGlyProProThGlyLeuProPr 432
 151CGCCACTTCTACTGCTTGTCTTCCAG 121
 |||:|||||
 432 oGlyProProProGlyAlaProProPhLeuAArgProProGlyMetProG 449
 120 GGTTCGATCTCCCTCTGCGGACTGTAAACCATAGATTGGCTATG 71
 |||: :|||
 449 lYLeuAArgGlyProLeuProAArg..... 456
 70 AGAGCCTGACTCTTATCTCTCTACCTCTTGACACCCCA 29
 |||||
 457LeuLeuProProGlyProProPro 464
 seq_name: /SIDB8/gcgdata/geneseq/geneseq/AA2001.DAT:AAV72165
 seq_documentation_block:
 ID AAV72165 standard; Protein: 641 AA.
 XX
 AC AAV72165;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human RNA metabolism protein (RMEP-5).
 XX
 KW Human: RNA metabolism protein; RMEP; nootropic; neuroleptic; antiulcer;
 KW transquilliser; antianemic; antidiabetic; immunosuppressive; cytostatic;
 KW antiaesthetic; antiinflammatory; anti-HIV; human immunodeficiency virus;
 KW antiarthritic; antiatherosclerotic; antiatherosclerotic; antiallergic;
 KW antipneumotoid; antiparkinsonian; antithyroid; nephrotrophic; antiout;
 KW thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;
 KW nervous system disorder; Alzheimer's disease; therapy; gene therapy;
 KW neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;
 KW peripheral nervous system disorder; cirrhosis; cranial nerve disorder;
 KW progressive neural autonomic nervous system disorder; Addison's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; drug screening.
 XX
 OS Homo sapiens.
 XX
 PN WO200078952-A2.

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT: AAB53462

seq_documentation_block:

ID AAB53462 standard; Protein; 647 AA.

AC AAB53462;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1002.

KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerrary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

PN W0200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05883.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI: 2000-587534/55.

DR N-PSDB; AAC98219.

PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -

PS Claim 11: Page 1580-1583; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53334 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal;
CC vulnerrary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders, immune
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 647 AA;

alignment_scores:

Quality: 68.50 Length: 64
Ratio: 2.635 Gaps: 2
Percent Similarity: 40.625 Percent Identity: 32.812

alignment_block:

US-09-823-101-2/rev x AAB53462 ..

Align seg 1/1 to: AAB53462 from: 1 to: 647

190 CCTCCGATGTCGTACCCACCTTTCCTCCCACT..... 152
||||| :||| |||||
422 ProProAlaProProLeuArgProProGlyProProThrGlyLeuProPr 438
151CGCCACCTTCTACTGCTTCTCTTCACG 121
438 oglyProProProGlyAlaProProPheLeuArgProProGlyMetProG 455
120 GCTTGGATCTCCCTCCCTCCGACGTGACCATGATTTGGCTATG 71
455 llyLeuArgGlyProLeuProArg..... 462
70 AGACCTGACTCTTATCTCTCTACTCTTGACACCCCA 29
||||| :|||
463LeuLeuProProGlyProProPro 470

seq_name: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT: AAM66399

seq_documentation_block:

ID AAM66399 standard; peptide; 42 AA.

AC AAM66399;

DT 12-JAN-1999 (first entry)

DE Cationic peptide Bac5.

KW Indolicidin analogue; resistance; cationic peptide; antibiotic;
KW bacterial infection; tolerance; antibacterial; microorganism;
KW bacteria; fungus; parasite; virus.

OS Bos taurus.

PN W09840401-A2.

PD 17-SEP-1998.

PF 10-MAR-1998; 98WO-CA00190.

PR 25-FEB-1998; 98US-0030619.

PR 10-MAR-1997; 97US-0040649.

PR 20-AUG-1997; 97US-0915314.

PR 26-SEP-1997; 97US-0060099.

PA (MICR-) MICROLOGIX BIOTECH INC.

PI Fraser JR, McNicol PJ, West MHP;

DR WPI: 1998-520800/44.

PT New indolicidin peptide analogues - useful for, e.g. enhancing
PT activity of antibiotic or overcoming tolerance, acquired resistance
PT or inherent resistance of microorganisms

PS Disclosure: Page 7; 105pp; English.

CC AAM66393 to AAM66469 represent native cationic peptides from the
CC present invention. The present invention describes compositions and
CC methods for treating infection, especially bacterial infections. The
CC compositions and methods use cationic peptides in combination with an
CC antibiotic agent which are then administered to a patient to enhance the
CC activity of the antibiotic agent, to overcome: (a) tolerance; (b)
CC acquired resistance; and (c) inherent resistance. The combinations of
CC antibiotics and cationic peptides can provide synergistic activity
CC against a microorganism that is tolerant, inherently resistant, or has
CC acquired resistance to an antibiotic agent. They can be used for killing
CC e.g. bacteria, fungi, parasites and viruses.

XX Sequence 42 AA;

alignment_scores:

Quality: 68.00 Length: 58

CC proliferation induction.
XX
SQ Sequence 43 AA;

alignment_scores:
Quality: 68.00 Length: 58
Ratio: 2.833 Gaps: 2
Percent Similarity: 41.379 Percent Identity: 27.586

alignment_block:
US-09-823-101-2/rev x AAR94447 ..

Align seg 1/1 to: AAR94447 from: 1 to: 43

```
199 GTCAGACACCCCTCCCTGATCTGTCACCCCAACCTTTCCTGCCCTCCACTCG 150
   ::|||:::|||||   ::   |||||   |||||   ||
6  IleArgArgProPro.....IleArgProProPheTyrProProPheAr 20
   |||   ::   ||   ::|||
149 CCACCTTCTACTGCGCTTGCTCTTCCAGGGTTTGATCTGCCCTTCCTC 100
   |||   ::   ||   ::|||
20  gProProIleArgProProIlePhePro..... 29
   |||   ::   ||   ::|||
99  GGGACTCTAGCCCATAGTATTTGGCTATGAGAGCGTACTCTTATCTCTC 50
   |||   ::   ||   ::|||
30  .....ProIle 31
   |||   ::   ||   ::|||
49  CTACCTCCTTGAGACACCCCACTC 26
   |||   ::   |||
32  ArgProProPheArgProProIleu 39
```


174 ...GTGACAGATCAGGAGG 191
::: |||||
326 SerAlaGlyGluArgGlyGly 332

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-09-214-564A-4

seq_documentation_block:

; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; TITLE OF INVENTION: Elongation By HIV-1 TAT
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

alignment_scores:

Quality:	66.00	Length:	57
Ratio:	1.941	Gaps:	2
Percent Similarity:	59.649	Percent Identity:	36.842

alignment_block:

US-09-823-101-2 x US-09-214-564A-4 ..

Align seg 1/1 to: US-09-214-564A-4 from: 1 to: 656

27 AGTGGGGGTGTCACAGAGGTAGAGATTAAGAGTCAGGCTCTCATAGC 76
||||||| ||||| ::|||
282 SerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSer..... 255
77 CAATACTATGCTTACAGTCCCGAGAGAGGGAGATCCAAACCTGGAA 126
||| ::||| |||||
296GlyProAspAsnArgGlyArgGlyArgGlyPheAspA 309
127 GAGACAGAGCAGTAGAGTGGCGAGTGGGAGGACGAGAAAGTTGG... 173
|| ||||| ::||| |||||
309 rglGlyGlyMetSerArgGlyGlyArgGlyGlyGlyArgGlyGlyMetGly 325
174 ...GTGACAGATCAGGAGG 191
::: |||||
326 SerAlaGlyGluArgGlyGly 332

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-349-696-25

seq_documentation_block:

; Sequence 25, Application US/08349696
; Patent No. 5599671
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.

STREET: P.O. Box 2000
CITY: Rahway
STATE: NJ
COUNTRY: United States
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh IIcx
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word 5.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/349,696

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: us/08/005945

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Meredith, Roy D.

REGISTRATION NUMBER: 30,777

REFERENCE/DOCKET NUMBER: 186991A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908)594-4678

TELEFAX: (908)594-4720

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 318 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-349-696-25

alignment_scores:

Quality:	65.50	Length:	54
Ratio:	2.113	Gaps:	2
Percent Similarity:	57.407	Percent Identity:	31.481

alignment_block:

US-09-823-101-2/rev x US-08-349-696-25 ..

Align seg 1/1 to: US-08-349-696-25 from: 1 to: 318

168 CTTTCGCGTCCCGCCAGCCACTCTACTGCTGTCTTCACAGG 119
||||| ||||| |||||
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleTrpAsnGly 257
118 TTGGATCTCCCTCTCGGAGCTGTAGCCATGATTGGCTATGAG 69
| ::||| |||
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCGTAGCTTATCTCTCTACTCTCTTGAGACACCCCACTACCAAG 19
||| ::||| |||||
265TyrMetGlyIleLeuLeuSerHisIleAsnSerMet 277
18 CAGCAGTGGTG 7
::||| ::|||
278 AsnProIleVal 281

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-233-009-25

seq_documentation_block:

; Sequence 25, Application US/08233009
; Patent No. 5646156
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: INHIBITION OF EOSINOPHIL
; TITLE OF INVENTION: ACTIVATION THROUGH A3 ADENOSINE RECEPTOR ANTAGONISM
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07065
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,009
; FILING DATE: 25-Apr-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Bence, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: 19219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3901
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
;
; US-08-233-009-25
```

```
alignment_scores:
  Quality: 65.50      Length: 54
  Ratio: 2.113      Gaps: 2
  Percent Similarity: 57.407      Percent Identity: 31.481
```

alignment_block:
US-09-823-101-2/rev x US-08-233-009-25 ..

Align seg 1/1 to: US-08-233-009-25 from: 1 to: 318

```
168 CTTTCCTGCTCCACATGCGCCACTTCTACTGCTGCTCTTCCAGG 119
||||| |||||||: :|||:|||||:|||||
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleTyrrPheAsnG1 257
118 TTGGAGTCTCCCTCTCTCGGAGCTGTAAAGCATAGATTGGCTATGAG 69
| :|||:| |
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCCTGACTTATATCTCTACTACTCTTGAGACACCCCACTACCAAGT 19
|||||: |||||||:|||||:|||||
265 .....TyrMetGlyIleLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCCAGTGTG 7
:::|||||:|
278 AsnProIleVal 281
```

seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-560-231-25

seq_documentation_block:

```
; Sequence 25, Application US/08560231
; Patent No. 5817760
;
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000
; CITY: Rahway
; STATE: NJ
; COUNTRY: United States
; ZIP: 07065
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh IIcx
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 5.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,231
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Meredith, Roy D.
; REGISTRATION NUMBER: 30,777
; REFERENCE/DOCKET NUMBER: 186991A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-4678
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-560-231-25
```

```
alignment_scores:
  Quality: 65.50      Length: 54
  Ratio: 2.113      Gaps: 2
  Percent Similarity: 57.407      Percent Identity: 31.481
```

alignment_block:
US-09-823-101-2/rev x US-08-560-231-25 ..

Align seg 1/1 to: US-08-560-231-25 from: 1 to: 318

```
168 CTTTCCTGCTCCACATGCGCCACTTCTACTGCTGCTCTTCCAGG 119
||||| |||||||: :|||:|||||:|||||
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleTyrrPheAsnG1 257
118 TTGGAGTCTCCCTCTCTCGGAGCTGTAAAGCATAGATTGGCTATGAG 69
| :|||:| |
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCCTGACTTATATCTCTACTACTCTTGAGACACCCCACTACCAAGT 19
|||||: |||||||:|||||:|||||
265 .....TyrMetGlyIleLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCCAGTGTG 7
:::|||||:|
278 AsnProIleVal 281
```

seq_name: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-080-704A-25

seq_documentation_block:

```
; Sequence 25, Application US/09080704A
; Patent No. 6166181
;
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000
```

CITY: Rahway
STATE: NJ
COUNTRY: United States
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT
SOFTWARE: Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,704A
FILING DATE: 18 May 1998
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18699DB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (732)594-4958
TELEFAX: (732)594-4720
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-080-704A-25

alignment_scores:
Quality: 65.50 Length: 54
Ratio: 2.113 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:
US-09-823-101-2/rev x US-09-080-704A-25 ..

Align seg 1/1 to: US-09-080-704A-25 from: 1 to: 318

```
168 CTTTCCTGCTCCGACCTCCGACCTCTCTGCTCTCTCCAGG 119
||||| |||||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
241 LeuSerTrpLeuProLeuSerIleIleAsnCySyleIleTyRheAsnGl 257
118 TTGGATCTCCCTTCCTCCGACGTAAAGCCATGATTGGCTATGAG 69
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCCTGACTCTTATCTCTCTACCTCTTGAGACACCCCACTCAGCAAGT 19
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
265 .....TyrMetGlyIleLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCCAGTGCTG 7
278 AsnProIleVal 281
```

seq_name: /cgn2_6/plodata/2/1aa/5B_COMB.pep:US-08-343-443B-107

seq_documentation_block:

; Sequence 107, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougaestel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS

TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; TITLE OF INVENTION: TRANSLATIONS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-107

alignment_scores:
Quality: 65.00 Length: 48
Ratio: 2.241 Gaps: 1
Percent Similarity: 60.417 Percent Identity: 39.583

alignment_block:
US-09-823-101-2 x US-08-343-443B-107 ..

Align seg 1/1 to: US-08-343-443B-107 from: 1 to: 25

```
27 AGTGGGGGTGTCCAGAGGTAGAGAGATAGAGATCAGGCTC 76
|||||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
46 SerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSer ... 59
77 CAATATGATGCTTACAGTCCCGAGAGAGGAGATCCAAACCCCTGGAA 126
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
60 .....GlyProAspAsnArgGlyArgGlyArgGlyArgGlyPheAspA 73
127 GAAGACAAGCACTAGAGAGTGGCCGAGTGGAGGCGACGAAAGGT 170
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
73 rgGlyGlyMetSerArgGlyGlyArgGlyGlyArgGlyGly 87
```

seq_name: /cgn2_6/plodata/2/1aa/6B_COMB.pep:US-09-540-245A-15

seq_documentation_block:

; Sequence 15, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc

```

; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-540-245A-15

alignment_scores:
      Quality: 64.00      Length: 64
      Ratio: 1.882      Gaps: 4
      Percent Similarity: 53.125      Percent Identity: 34.375

alignment_block:
US-09-823-101-2/rev x US-09-540-245A-15 ..

Align seq 1/1 to: US-09-540-245A-15 from: 1 to: 1395

176 CACCACCTTTCCTCCCTCCACCTG.....CCCACTTCTACTGC 136
||||| ||| ::::::::::: |||||||
2 HisPromeHisProGluAsnHisAlaIleAlaArgSerThrSerThp 18
135 CTGTCT.....TCTTCAGGCTTGGATCTCCCT.... 105
: :: ||||||| :::: |||
18 rAsnAsnProSerArgSerArgSerArgMetTrpLeuProLat 35
104 .....TCTCGGAGCTGTAGCCATAGTATT 78
||||:||||| ||:::
35 rpleuLeuValLeuValAlaSerAsnGlyLeuProAlaValArgGly 51
77 GGCATAGAGCCCTGACTTATCTCT..CTTACCTCCCTG 39
||||:||||| :::: |||
52 GlnTyrGlnSerProArgIleIleGlnHisProThrAspLeu 65

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-362-670B-30

seq_documentation_block:
; Sequence 30, Application US/08362670B
; Patent No. 5658862
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Moznay, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolman, Neil A.
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,670B
; FILING DATE: December 22, 1994
; CLASSIFICATION: 514
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-670B-30

alignment_scores:
      Quality: 63.50      Length: 63
      Ratio: 1.924      Gaps: 1
      Percent Similarity: 52.381      Percent Identity: 26.984

alignment_block:
US-09-823-101-2 x US-08-362-670B-30 ..

Align seq 1/1 to: US-08-362-670B-30 from: 1 to: 240

30 GGGGTGTCAGAGAGTAGAGATAGATAGATGAGCTCTCAT..... 74
||||| ||||| ::::::::::: |||||||
29 GlnAlaAspGlyGlyGlyThrAla***GluGlnArgAlaLeuValI 45
75 .....GCCA 78
|||:
45 eSerSerArgThrGlnArgGlySerLeuPheArgGluIleArgAlaG 62
::: ||::: |||
79 AATCATGCTGCTTACAGTCCCGAGAGAGGAGAGATCCAAACCTGGA 128
||||:||||| :::: |||
62 IlnAlaArgAlaLeuArgAlaAlaIleAlaGluProProAspProGlyPro 78
129 AGACAAGGCAGTAGAGTGGGAGGAGGAGGAGGAGAA 167
||||| :::: |||
79 GlyAlaGlySerArgGlyAlaAsnLeuGlyGlyArgArg 91

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-333-576C-30

seq_documentation_block:
; Sequence 30, Application US/08333576C
; Patent No. 6027919
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Moznay, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolman, Neil A.
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/333,576C
; FILING DATE: No. 6027919ember 2, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-333-576C-30

alignment_scores:
Quality: 63.50 Length: 63
Ratio: 1.924 Gaps: 1
Percent Similarity: 52.381 Percent Identity: 26.984

alignment_block:
US-09-823-101-2 x US-08-333-576C-30 ..

Align seg 1/1 to: US-08-333-576C-30 from: 1 to: 240

30 GGGGGTGTCCAGAGAGTACAGATAGAGTCTCATATA..... 74
||||| ||||| :|||||:|||||:
29 GYGlyAapGlyGlyGlyThrAla**GlUGluArGAlaLeuLeuVal11 45
75GCCA 78
45 eSerSerArGThrGlnArGlyGluSerLeuPheArGluIleArGAlaG 62
79 AATACTATGCTTACATCCCGAGAGGAGATCCAAACCTCGAGA 128
:::|||||:|||||:|||||:|||||:
62 InAlArGAlaLeuArGAlaAlaAlaGluProProPtoAsPProGlyPro 78
129 AGACAAGCAGTAGAAGTGGCGAGTGGAGGAGCAGGA 167
|||||:|||||:|||||:|||||:
79 GYAlaGlySerArGlyAlaAlaAlaAlaAlaGlyGlyArGArg 91

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-808-324-30

seq_documentation_block:

; Sequence 30, Application US/08808324
; Patent No. 6284872
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Wozney, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolfman, Neil
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,324
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.

REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-324-30

alignment_scores:
Quality: 63.50 Length: 63
Ratio: 1.924 Gaps: 1
Percent Similarity: 52.381 Percent Identity: 26.984

alignment_block:
US-09-823-101-2 x US-08-808-324-30 ..

Align seg 1/1 to: US-08-808-324-30 from: 1 to: 240

30 GGGGGTGTCCAGAGAGTACAGATAGAGTCTCATATA..... 74
||||| ||||| :|||||:|||||:
29 GYGlyAapGlyGlyGlyThrAla**GlUGluArGAlaLeuLeuVal11 45
75GCCA 78
45 eSerSerArGThrGlnArGlyGluSerLeuPheArGluIleArGAlaG 62
79 AATACTATGCTTACATCCCGAGAGGAGATCCAAACCTCGAGA 128
:::|||||:|||||:|||||:|||||:
62 InAlArGAlaLeuArGAlaAlaAlaGluProProPtoAsPProGlyPro 78
129 AGACAAGCAGTAGAAGTGGCGAGTGGAGGAGCAGGA 167
|||||:|||||:|||||:|||||:
79 GYAlaGlySerArGlyAlaAlaAlaAlaAlaGlyGlyArGArg 91

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US94-14030A-30

seq_documentation_block:

; Sequence 30, Application PC/TUS9414030A
; GENERAL INFORMATION:
; APPLICANT: GENETICS INSTITUTE, INC.
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14030A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/164,103
; FILING DATE: 07-DEC-1993
; APPLICATION NUMBER: US 08/217,780
; FILING DATE: 25-MAR-1994
; APPLICATION NUMBER: US 08/333,576
; FILING DATE: 02-NOV-1994
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202D-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-14030A-30

```

```

alignment_scores:
  Quality: 63.50      Length: 63
  Ratio: 1.924      Gaps: 1
  Percent Similarity: 52.381      Percent Identity: 26.984

```

alignment_block:

US-09-823-101-2 x PCT-US94-14030A-30 ..

Align seg 1/1 to: PCT-US94-14030A-30 from: 1 to: 240

```

30 GGGGCTGCCAAGAGGAGATAGATCAGGCTCTCAT..... 74
||||| ||||| :|||||
29 GlyIyspArgIyGlyThAla**GluluarGAlaLeuValI 45
75 .....GCCA 78
45 eSerSerArgThrGlnArgIyGlySerLeuPheargIuIleargIag 62
79 AATACATGGCTTACAGTCCCGAGAGGAGATCCAAACCTCGAAGA 128
:||||| :||||| :|||||
62 InAlaArgAlaLeuArgAlaAlaIaGlupProAspProGIyArg 78
129 AGACAAGGACAGTACAGTGGCGGAGTGAGGAGCAAGAA 167
||||| :||||| :|||||
79 GlyIaGlySerArgIyAlaAsnLeuGIyGIyArgArg 91

```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-897-924A-8

seq_documentation_block:

```

; Sequence 8, Application US/08897924A
; Patent No. 6028058
; GENERAL INFORMATION:
; APPLICANT: Flokiewicz, Robert Z.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING
; TITLE OF INVENTION: NUCLEAR TRAFFICKING OF PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,924A
; FILING DATE: 21-JUL-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200124.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

```

```

; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-897-924A-8

```

```

alignment_scores:
  Quality: 62.50      Length: 30
  Ratio: 3.289      Gaps: 1
  Percent Similarity: 63.333      Percent Identity: 53.333

```

alignment_block:

US-09-823-101-2 x US-08-897-924A-8 ..

Align seg 1/1 to: US-08-897-924A-8 from: 1 to: 67

```

90 TTACAGTCCGAGGAGGAGATCCAAACCTGGAAGACACAGGCGAG 139
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1 LeuGIyAspArgIyArgGIyArgAlaLeuProGIyGIyArgLeuGIyGI 17
140 TAGAAGT.....GGCGAGTGGGAGGCGAGG 164
:||||| :||||| :||||| :||||| :|||||
17 YArgGIyArgGIyArgAlaProGIyArgValGIyGIyArg 30

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-2 to: PIR-68.* out_format: pfs
 Date: Jan 17, 2002 4:03 PM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlP
 -O=/cgn2.1/USPRO.spool/US09823101/runat.17012002.155137_22030/app_query.fasta_1.2239
 -DB=PIR-68 -OFMT=fastan -SUFFIX=n2p.rpr -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsom62
 -TRANS=human40.cdi -LIST=45 -DOCALLN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORMEXT -MINLEN=0 -MAXLEN=200000000
 -USER=US09823101@cgn1_1.200 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
 -WAIT -THEADS=1

Search information block:

Query: US-09-823-101-2
 Query Length: 235
 Database: PIR-68.*
 Database sequences: 219241
 Database length: 76174552
 Search time (sec): 120.350000

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
PIR2:T33997	75.50	140.62	1.63	471	hypothetical protein W03G1.5 -
PIR2:S72880	75.00	143.34	1.82	298	hypothetical protein B2126.F3.1
PIR2:T15667	75.00	138.38	1.87	545	hypothetical protein C27H5.3 -
PIR2:J01663	71.50	136.01	4.60	301	hybrid proline-rich protein, p
PIR2:S04536	70.50	139.28	5.80	157	embryonic abundant protein, hom
PIR2:A58768	70.50	127.48	6.25	661	ATP-dependent RNA helicase, hom
PIR2:T21689	70.00	135.93	6.72	208	hypothetical protein F33A8.3 -
PIR2:JC4817	70.00	135.90	6.72	209	RNA-binding protein R2.1 - wood
PIR2:T16578	69.50	134.40	7.70	221	hypothetical protein K07E1.1 -
PIR2:T33251	68.50	122.95	10.66	693	hypothetical protein C18G1.4 -
PIR2:A45328	68.00	144.92	10.51	42	bactenecin 5 - bovine
PIR2:P6687	68.00	133.16	11.33	176	bactenecin 5 precursor - bovine
PIR2:E96576	67.50	129.30	11.62	282	hypothetical protein T6119.1 [1
PIR2:A36502	67.50	115.72	14.37	523	unknown protein, 43598-45751 [1
PIR2:S41773	67.00	131.62	14.73	1300	insulin receptor-related recep
PIR2:S59529	67.00	131.62	14.73	165	glycine-rich RNA-binding protei
PIR2:T09555	67.00	126.19	15.25	320	fibrillarin - Arabidopsis thali
PIR2:S06028	67.00	117.14	16.16	964	gene suppressor-of-white-apric
PIR2:S33690	66.50	125.55	17.37	305	fibrillarin - fission yeast (Sc
PIR2:NMBE1	66.50	124.65	17.47	340	latency-related protein 1 - hum
PIR2:T33022	66.50	117.65	18.27	798	hypothetical protein K07H8.10 -
PIR2:TS1029	66.00	114.29	18.67	1203	related to pathway-specific n
PIR2:G84846	66.00	125.07	19.77	285	40S ribosomal protein S2 [imp
PIR2:A49358	66.00	118.23	20.65	656	RNA-binding protein EMS - human
PIR2:JC7174	66.00	117.00	20.82	762	N,N-dimethylornithinase (EC 3.5
PIR2:S38511	65.50	123.13	22.70	318	adenosine receptor A3 - human
PIR2:S2421	65.50	122.90	22.74	327	nucleolar protein NOP1 - yeast
PIR2:TS3174	65.00	122.71	25.83	295	RNA binding protein/transcript
PIR2:TS6147	65.00	121.97	25.95	323	sulfotransferase-like protein -
PIR2:FF6537	65.00	121.84	25.97	328	A/G-specific adenine glycosylas
PIR2:S74790	65.00	120.38	26.21	392	hypothetical protein s111025 -
PIR2:T38951	65.00	112.46	27.58	1029	probable ubiquitin transferase
PIR2:T00047	64.50	105.26	28.88	2475	gelatin lyase (EC 4.2.2.-) prec
PIR2:T04126	64.50	128.22	28.89	133	preprotein translocase secA hom
PIR2:FOMVGC	64.50	119.80	29.85	371	gag polyprotein - feline sarcom
PIR2:T06458	64.50	118.68	30.06	425	gag polyprotein - feline sarcom
PIR2:TVWVR	64.50	115.70	30.64	611	nucleolin homolog - garden pea
PIR2:T46289	64.50	115.03	30.77	663	protein-tyrosine kinase (EC 2.7
		112.88	31.20	862	hypothetical protein DKF2p34A1

PIR2:T00705 - 64.50 112.50 31.28 903 i N-chimerin homolog F25965.3
 PIR2:T02367 + 64.50 107.38 32.32 1684 i hypothetical protein T8F5.2
 PIR2:T38091 - 64.50 106.65 32.47 1841 i cell division control prote
 PIR2:T16208 + 64.00 129.12 31.90 105 i hypothetical protein F30H5.2
 PIR2:J51417 + 64.00 119.90 33.84 323 i fibrillarin - African clawed

seq_name: PIR2:T33997

seq_documentation_block:

hypothetical protein W03G1.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33997
 R:Pauley, A.; Scheel, P.; Harper, M.
 submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid W03G1.
 A:Reference number: Z21454
 A:Accession: T33997
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-471 <PAU>
 A:Cross-references: EMBL:AF125964; PIDN:AA14753.1; GSPDB:GN00022; CESP:W03G1.5
 A:Experimental source: strain Bristol N2; clone W03G1
 C:Genetics:
 A:Gene: CESP:W03G1.5
 A:Map position: 4

alignment_scores:

Quality: 75.50 Length: 49
 Ratio: 2.517 Gaps: 2
 Percent Similarity: 61.224 Percent Identity: 42.857

alignment_block:

US-09-823-101-2 x T33997 ..

Align seg 1/1 to: T33997 from: 1 to: 471

39 GAAGAGCGTGAAGAGTAAAGTCAGCTGCATGACCCAAATACATG 88
 |||||:||||| ||| ||
 206 GAGGAGTGAAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 222
 |||||:||||| ||| ||
 89 CTTCAGTCCGAGGAGAGG...GGAGATCA.....AACCTG 123
 |||||:||||| ||| ||
 222 yvalgluhtsargluyargserglyserprogluyargatrgluyhsg 239
 |||||:||||| ||| ||
 124 GAAGAGACAGACAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 170
 |||||:||||| ||| ||
 239 lgluyarghtsargluyargserglyserprogluyarghtsargluy 254

seq_name: PIR2:S72880

seq_documentation_block:

hypothetical protein B2126.F3.142 - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
 C:Accession: S72880
 R:Smith, D.R.; Robison, K.
 submitted to the EMBL Data Library, November 1993
 A:Description: Mycobacterium leprae cosmid B2126.
 A:Reference number: S72585
 A:Accession: S72880
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-298 <SMI>
 A:Cross-references: EMBL:U00017; NID:9466994; PIDN:AA17220.1; PID:9467035
 C:Genetics:
 A:Start codon: GTG

alignment_scores:

Quality: 75.00 Length: 55
 Ratio: 2.143 Gaps: 1

Percent Similarity: 63.636 Percent Identity: 34.545

alignment_block:

US-09-823-101-2/rev x S72880 ..

Align seg 1/1 to: S72880 from: 1 to: 298

```

185 CTGATCTGTACACCACTTCCGCTCCGCCACACGCCCACTTCTACTGC 136
||||| : : : : : : : : : : : : : : : : : : : : : : :
212 LeuAlaLeuAaGlnTrnTrpValAlaRghIsaIaProTyrGlyLeuAl 228
||||| : : : : : : : : : : : : : : : : : : : : : : :
135 CTGTCTCTCTCCAGGTTTGATCTCCCTCCCGGAGACTGAAGCCA 86
| ||| : : : : : : : : : : : : : : : : : : : : : : :
228 aglySerArgThrLeuAlaTrpValSerGlnThrAlaSer.....Asn 243
||||| : : : : : : : : : : : : : : : : : : : : : : :
85 TAGTATTTGGCTATGAGAGCCCTTATCTCTCCACCTCGTTGAC 36
||||| : : : : : : : : : : : : : : : : : : : : : : :
243 leValIleGlyTyrAspLeuThrThrGlyIleProValGluLysValSer 259

35 ACCCCCACTCACCA 21
||||| |||
260 TyrProThrValGln 264

```

seq_name: p1r2:TI5667

seq_documentation_block:

hypothetical protein C27H5.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
 C:Accession: T15667
 R:Pauley, A.
 submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of C. elegans cosmid C27H5.
 A:Reference number: Z18386
 A:Accession: T15667
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-545 <PAU>
 A:Cross-references: EMBL:U14535; NID:9540265; PID:9540269; PIDN:AAC46657.1; CESP:C27H5.3
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:C27H5.3
 A:Introns: 38/2; 52/1; 87/3; 183/1; 202/2; 262/1; 308/2; 511/3
 C:Superfamily: RNA-binding protein EMS; ribonucleoprotein repeat homology

alignment_scores:

Quality:	75.00	Length:	65
Ratio:	2.419	Gaps:	3
Percent Similarity:	47.692	Percent Identity:	38.462

alignment_block:

US-09-823-101-2 x T15667 ..

Align seg 1/1 to: T15667 from: 1 to: 545

```

27 AGTGGGGTGTCCAGAGAGTGAAGATAGAGTCAAGCTCTCATAGC 76
||||| : : : : : : : : : : : : : : : : : : : : : : :
123 serGlyGlnserGly.....Se 130
||||| : : : : : : : : : : : : : : : : : : : : : : :
77 CAATACTATGGCTTACAGTCCCGAGAGGAGGATCCAAAC..... 119
||||| : : : : : : : : : : : : : : : : : : : : : : :
130 rAspProTyrGly...GlnserArgGlyGlyArgGlyGlyPheGly 146
||||| : : : : : : : : : : : : : : : : : : : : : : :
120 .....CTGGAAGAGACAGAGCAGCTAGAGT 146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 lySerArgGlyGlyGlyTyrAspGlyGlyArgGlyGlySerArgGly 162
147 GGGCGAGTGGAGAGCAAGGTTGGTGCACATCAGGAGG 191
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 GlyTyrAspGlyGlyArgGlyGlyTyrGlyGlyAlaPargGlyGly 177

```

seq_name: p1r2:JQ1663

seq_documentation_block:

hybrid proline-rich protein - maize
 C:Species: *zea mays* (maize)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Sep-1999
 C:Accession: JQ1663
 R:Jose-Estanyol, M.; Ruiz-Avila, L.; Pulidomench, P.
 Plant Cell 4, 413-423, 1992
 A:Title: A maize embryo-specific gene encodes a proline-rich and hydrophobic protein.
 A:Reference number: JQ1663; MUID:92361259
 A:Accession: JQ1663
 A:Molecule type: DNA
 A:Residues: 1-301 <OS>
 A:Cross-references: EMBL:X60432; NID:g433706; PIDN:CAA42959.1; PID:g433707
 A:Experimental source: strain W64A
 C:Superfamily: hydroxyproline-rich glycoprotein

alignment_scores:

Quality:	71.50	Length:	61
Ratio:	2.750 <td>Gaps:</td> <td>3 </td>	Gaps:	3
Percent Similarity:	42.623 <td>Percent Identity:</td> <td>29.508 </td>	Percent Identity:	29.508

alignment_block:

US-09-823-101-2/rev x JQ1663 ..

Align seg 1/1 to: JQ1663 from: 1 to: 301

```

190 CCCTCCGATCTGCACCAACTTTCCTGCTCCGACCTGCTTCT 141
||||| ||| ||| : : : : : : : : : : : : : : : : : :
161 ProPro.....ThrProTyrValProProTyrValProTyrVa 174
||||| : : : : : : : : : : : : : : : : : : : : : : :
140 ACTGCTCTGT.....CTCTCCAGGCTTGGAT 112
||||| : : : : : : : : : : : : : : : : : : : : : : :
174 lProProThrProArgProSerProProTyrValProProTyrValP 191
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 CTCCTCTCTCGGAGCTGAAGCCATGATTTGGCTATGAGAGCCTGA 62
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 roProThrProArgProSerProPro..... 199
61 CTCTATCTCTCCTACCTCTTGAGACCCCA 29
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 .....TyrValProProTyrValProPro 207

```

seq_name: p1r2:S04536

seq_documentation_block:

embryonic abundant protein, glycine-rich - maize
 C:Species: *zea mays* (maize)
 C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 23-Jul-1999
 C:Accession: S04536
 R:Gomez, J.; Sanchez-Martinez, D.; Stiefel, V.; Rigau, J.; Pulidomench, P.; Pages, M.
 Nature 334, 262-264, 1988
 A:Title: A gene induced by the plant hormone abscisic acid in response to water stress
 A:Reference number: S04536; MUID:88288401
 A:Accession: S04536
 A:Molecule type: DNA
 A:Residues: 1-157 <GOM>
 A:Cross-references: EMBL:X12564; NID:g22312; PIDN:CAA31077.1; PID:g22313
 C:Genetics:
 A:Introns: 38/3
 C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology
 F:9-76/Domain: ribonucleoprotein repeat homology <RM1>

alignment_scores:

Quality:	70.50	Length:	53
Ratio:	2.431 <td>Gaps:</td> <td>1 </td>	Gaps:	1
Percent Similarity:	54.717 <td>Percent Identity:</td> <td>37.736</td>	Percent Identity:	37.736

alignment_block:

US-09-823-101-2 x S04536 ..

Align seg 1/1 to: S04536 from: 1 to: 157

```

30 GGGGGTGTCCAGAGAGTAGAGATAGACTCAGCTCTCATAGCCAA 79
||||| ||||||||| ||| |||
103 GYGLYTYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGL 119
80 ATACTATGGCTTACAGTCCCGAGAGAGGAGATCCAAAC..... 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 YTYRGLYGLYARGARGGLUGLYGLYGLYGLYGLYGLYGLYGLY 136
120 ....CTCGAAGAGACAGCAGCTAGAGAGCGGCGAGTGGAGGAGC 164
||||| ||||||||| ||| ||| ||| ||| ||| ||| ||| |||
136 YTYRGLYGLYARGARGGLUGLYGLYGLYGLYGLYGLYGLYGLY 152
165 AAGGTTGG 173
|||||
153 GLYGLYTRP 155

```

seq_name: p1r2:A58768

seq_documentation_block:

ATP-dependent RNA helicase homolog - fruit fly (*Drosophila melanogaster*)

C/Species: *Drosophila melanogaster*

C/Date: 31-Dec-1990 #sequence_revision 17-Apr-1998 #text_change 19-Jan-2001

C/Accession: A58768; S01676; S10129; A31922

R/Liang, L.; Diehl-Jones, W.; Lasko, P.

unpublished results, 1995, cited by GenBank in release 103.0

A/Description: Localization of Vasa protein to the *Drosophila* pole plasm is independent

A/Reference number: A58768

A/Accession: A58768

A/Status: not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-661 <LASI>

A/Cross-references: EMBL:X12945; NID:9433675; PIDN:CAA31405.1; PID:91054723

R/Lasko, P.F.; Ashburner, M.

Nature 333, 611-617, 1988

A/Title: The product of the *Drosophila* gene vasa is very similar to eukaryotic initiator

A/Reference number: S01676; MUID:89014721

A/Accession: S01676

A/Molecule type: DNA

A/Residues: 1-34, 'R', 36-448, 'LRF', 453-460, 'R', 462-589, 'Q', 591-661 <LAS2>

A/Cross-references: EMBL:X12945; NID:9433675

R/Ashburner, M.

submitted to the EMBL Data Library, October 1988

A/Reference number: S10129

A/Accession: S10129

A/Molecule type: DNA

A/Residues: 1-34, 'R', 36-644, 'Q', 645-661 <ASH>

A/Cross-references: EMBL:X12945; NID:9433675

R/Hay, B.; Jan, L.Y.; Jan, Y.N.

Cell 55, 577-587, 1988

A/Title: A protein component of *Drosophila* polar granules is encoded by vasa and has ext

A/Reference number: A31922; MUID:89028669

A/Accession: A31922

A/Molecule type: mRNA

A/Residues: 1-34, 'R', 36-154, 168-264, 'Y', 266-321, 'C', 323-451, 'F', 453-581, 'R', 583-593, 'H',

A/Cross-references: GB:M23560; NID:9158795; PIDN:AAA29013.1; PID:9158796

C/Genetics:

A:Gene: vasa

A/Cross-references: FlyBase:FBgn0003970

A/Introns: 8/3; 138/1; 234/3; 482/3; 554/1; 644/2

C/Superfamily: ATP-dependent RNA helicase DBP1

C/Keywords: ATP; nucleotide binding; P-loop

F:288-296/Region: nucleotide-binding motif A (P-loop)

F:395-400/Region: nucleotide-binding motif B

F:399-402/Region: DEAD motif

Alignment_scores: Quality: 70.50 Length: 68

Percent Similarity: 47.059 Percent Identity: 33.824

alignment_block:

us-09-823-101-2 x A58768 ..

Align seg 1/1 to: A58768 from: 1 to: 661

```

30 GGGGGTGTCCAGAGAGTAGAGATAGACTCAGCTCTCATAGCCAA 79
||||| ||||||||| ||| |||
48 GYGLYTYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGL 56
80 ATACTATGGCTTACAGTCCCGAGAGAGGAGATCA..... 116
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 ValPheGlyArgIleGlyGlyArgGlyGlyGlyGlyGlyGly 73
117 .....AACCTGGAAGAACAAAGC 137
73 rgGlyGlyAsnArgAspGlyGlyGlyPheHisGlyGlyArgArgGly 89
138 AGTAGAGTGGCGAGTGGAGCAGAGAAAGTTGGTGACAGATCAGG 187
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 GluArgAspPheArg...GlyGlyGlyGlyGlyGlyPheArgGlyGly 105
188 AGCG 191
|||||
105 yGly 106

```

seq_name: p1r2:T21689

seq_documentation_block:

hypothetical protein F33A8.3 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C/Accession: T21689

R/Mathews, L.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z19459

A/Accession: T21689

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-208 <WILL>

A/Cross-references: EMBL:Z81525; PIDN:CA804257.1; GSPDB:GN00020; CESP:F33A8.3

A/Experimental source: clone F33A8

C/Genetics:

A:Gene: CESP:F33A8.3

A:Map position: 2

A/Introns: 5/3; 174/3

C/Superfamily: Arabidopsis glycine-rich protein 2; cold shock domain homology

F:23-87/Domain: cold shock domain homology <CSD>

alignment_scores: Quality: 70.00 Length: 77

Percent Similarity: 48.052 Percent Identity: 35.065

Alignment_block:

us-09-823-101-2 x T21689 ..

Align seg 1/1 to: T21689 from: 1 to: 208

```

30 GGGGGTGTCCAGAGAGTAGG.....AGAGATAGAGTCAGGC 67
||||| ||||||||| ||| |||
93 GlyProValInclInclSerIstYrAlaAlaAspArgAspAlaGlu 109
68 TCTCATAGCCAAATACTATGCTTACAGTCCGAGAGAAG..... 107
109 a.....AlaArgGlyArgGlyGlyArgGlyArgGlyArgGly 124
108 .....GGA 110
124 rgGlyGlyIleArgHisAspSerGlySerArgAspAlaGluGlyGly 140
111 GATCCAAACCTTGGAAGAGACAAAGCAGTGAAGTGGCGAGTGGAG 160

```


THIS PAGE BLANK (USPTO)


```

Align seq 1/1 to: GRPA_MAIZE from: 1 to: 157
30 GGGGGTGTCCAGAGGAGTAGAGATAGAGTCAGGCTCATAGCCAA 79
||||| ||||||||| |||
103 GYGLYTYRGLYGLYGLYARGARGASDGLYGLYTYRGLYGLYGLYGL 119
80 ATACTATGCTTACAGTCCCGAGGAGGAGATCAAC..... 119
||| ||| ||| ||| |||
119 YTYRGLYGLYARGARGGLUGLYGLYGLYGLYTYRGLYGLYGLYGL 136
120 .....CCTGGAGAGACAGCAGTAGAAGTGGCCAGTGGGAGGAGG 164
||||| ||||||| |||
136 LYTYRGLYGLYARGARGGLUGLYGLYGLYGLYTYRGLYGLYGLY 152
165 AAAGGTGG 173
|||||
153 GYGLYTRP 155

seq_name: SwissProt_39:VASA_DROME

seq_documentation_block:
ID VASA_DROME STANDARD: PRT: 661 AA.
AC P09052: Q24582: Q9V308;
DT 01-NOV-1988 (Rel. 09, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VASA PROTEIN (ANTIGEN MAB46F11).
GN VAS OR BG:DS00929.14 OR CG3506
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89014721; PubMed=3140040;
RA Laslo P.F.; Ashburner M.;
RT "The product of the Drosophila gene vasa is very similar to
RT eukaryotic initiation factor-4A."
RL Nature 335:611-617(1988).
RN [2]
RP REVISIONS.
RA Laslo P.F.;
RL Submitted (DEC-1993) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89028659; PubMed=3052853;
RA Hay B.; Jan L.Y.; Jan Y.N.;
RT "A protein component of Drosophila polar granules is encoded by vasa
RT and has extensive sequence similarity to ATP-dependent helicases."
RL Cell 55:577-587(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-BERKELEY;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M.; Misra S.; Roote J.; Lewis S.E.; Blazej R.; Davis T.;
RA Doyle C.; Galle R.; George R.; Harris N.; Hartzell G.; Harvey D.;
RA Hong L.; Houston K.; Hoskins R.; Johnson G.; Martin C.; Moshrefi A.;
RA Palazolo M.; Reese M.G.; Spradling A.; Tsang G.; Wan K.; Whitefaw K.;
RA Celniker S.; Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region."
RL Genetics 153:179-219(1999).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D.; Celniker S.E.; Holt R.A.; Evans C.A.; Gocayne J.D.;
RA Amanatides P.G.; Scherer S.E.; Li P.W.; Hoskins R.A.; Galle R.F.;
RA George R.A.; Lewis S.E.; Richards S.; Ashburner M.; Henderson S.N.;
RA Sutton G.G.; Wortman J.R.; Yandell M.D.; Zhang Q.; Chen L.X.;
RA Brandon R.C.; Rogers Y.-H.C.; Blazej R.G.; Champe M.; Pfeiffer B.D.;
RA Wan K.H.; Doyle C.; Baxter E.G.; Helt G.; Nelson C.R.; Miklos G.L.G.;

```

```

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: THE VASA PROTEIN IS REQUIRED ONLY IN THE FEMALE GERM
CC LINE. IT IS IMPORTANT FOR OOCYTE FORMATION AND IN THE
CC SPECIFICATION OF THE POSTERIOR STRUCTURES OF THE EMBryo.
CC -!- DEVELOPMENTAL STAGE: MATERNALLY EXPRESSED (DURING OOGENESIS).
CC FUNCTION DURING EARLY EMBRYOGENESIS.
CC -!- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE.
CC DD4/VASA SUBFAMILY
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
CC -----
CC EMBL: X12945; CAA31405.1; -
CC EMBL: X12946; CAA31405.1; JOINED.
CC EMBL: M23560; AAA29013.1; -
CC EMBL: AE003412; AAF44917.1; -
CC EMBL: AE003646; AAF53438.1; -
CC PIR: A31922; A31922.
CC PIR: S01676; S01676.
CC FLYbase: FBgn003970; vas.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR000629; DEAD_ATP_helicase.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF00270; DEAD. 1.
CC Pfam: PF00271; Helicase_C. 1.
CC SMART: SM00487; HXDC. 1.
CC SMART: SM00490; HELIC_C. 1.
CC PROSITE: PS00039; DEAD_ATP_HELICASE. 1.
CC Developmental protein; ATP-binding; Repeat; Helicase.
CC NP_BIND 289 296 ATP (BY SIMILARITY).
CC SITE 399 402 DEAD BOX.
CC DOMAIN 93 127
CC REPEAT 93 99 [EQ]-G-G.
CC REPEAT 100 106 1.
CC REPEAT 107 113 2.
CC REPEAT 114 120 3.
CC REPEAT 114 120 4.

```

```

RP  [3] STRUCTURE BY NMR OF 131-173.
RX  MEDLINE=96183292; PubMed=8605180;
RA  Raj P.A., Marcus E., Edgerton M.;
RT  "Delineation of an active fragment and poly(L-proline) II
RT  conformation for candidal activity of bacitracin 5.";
RN  Biochemistry 35:4314-4325(1996).
RN  [4]
RP  CHARACTERIZATION.
RX  MEDLINE=96300243; PubMed=8706679;
RA  Storici P., Tossi A., Lenarcic B., Romeo D.;
RT  "Purification and structural characterization of bovine
RT  cathelicidins, precursors of antimicrobial peptides.";
RL  Eur. J. Biochem 238:769-776(1996)
CC  -1- FUNCTION: EXERTS, IN VITRO, A POTENT ANTIMICROBIAL ACTIVITY.
CC  PROBABLY DUE TO AN IMPAIRMENT OF THE FUNCTION OF THE RESPIRATORY
CC  CHAIN AND OF ENERGY-DEPENDENT ACTIVITIES IN THE INNER MEMBRANE
CC  OF SUSCEPTIBLE MICROORGANISMS.
CC  -1- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.
CC  -1- DOMAIN: BACS SEQUENCE CONSISTS ALMOST EXCLUSIVELY OF X-P-P-Y
CC  REPEATS.
CC  -1- PTM: ELASTASE IS RESPONSIBLE FOR ITS MATURATION.
CC  -1- MASS SPECTROMETRY: MW=16570; MW_ERR=1; METHOD=ELECTROSPRAY;
CC  RANGE=30-173.
CC  -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L02650; AAA30404.1; .
DR  PIR; B36589; B36589.
DR  PIR; A45328; A45328.
DR  InterPro; IPR001894; Cathelicidin.
DR  InterPro; IPR000010; Cystatin.
DR  Pfam; PF00666; Cathelicidins; 1.
DR  ProDom; PD001838; Cathelicidin; 1.
DR  SMART; SM00043; Cy; 1.
DR  PROSITE; PS00946; CATHELICIDINS_1; 1.
DR  PROSITE; PS00947; CATHELICIDINS_2; 1.
KW  Antibiotic; Repeat; signal; Amidation.
FT  SIGNAL 1 29 POTENTIAL.
FT  PROPEP 30 130
FT  PEPTIDE 131 173 BACTENECIN 5.
FT  PROPEP 174 176 REMOVED IN MATURE FORM.
FT  MOD_RRS 30 30 PYROLIDONE CARBOXYLIC ACID.
FT  DISULFID 85 96 BY SIMILARITY.
FT  DISULFID 107 124 BY SIMILARITY.
FT  MOD_RRS 173 173 AMIDATION (G-174 PROVIDE AMIDE GROUP).
FT  CONFLICT 170 171 GP -> R (IN REF. 2).
SQ  SEQUENCE 176 AA; 20030 MW; D2D1881C05929415 GPC64;

alignment_scores:
    quality: 68.00      length: 58
    ratio: 2.833        gaps: 2
    percent similarity: 41.379    percent identity: 27.586

alignment_block:
US-09-823-101-2/rev x BCTS_BOVIN
Align seg 1/1 to: BCTS_BOVIN from: 1 to: 176

199  GTGACACACCCCTGCTGCATCTGCACCCAACTTCCTGCTCCACTCG 150
    :::::::::::::::  ::  ::::::::::  ::
136  ILeatgAgtProPro.....ILeatgPProPhetYrProProhear 150
149  CCCACTTACTGCTGTCTTCTTCACAGGTTTGAGATCCCTCCCTTCGTC 100
    ||||  ::  |||  ::::::::::|||

```

```

150 gProBrolleArgProBrollePhePro..... 159
99 GGGAGCTGTAAGCCATGATATTGGCTAGAGAGCCGTGACTCTATCTCTC 50
160 .....Protle 161
49 CTACCTCTTGACACCCCACTC 26
162 ArgProBrolleArgProBrolle 169

seq_name: SwissProt_39:IRR_CAVPO

seq_documentation_block:
ID IRR_CAVPO STANDARD; PRT; 1300 AA.
AC P14617;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE INSULIN RECEPTOR-RELATED PROTEIN PRECURSOR (EC 2.7.1.112) (IRR)
DE (IRR-RELATED RECEPTOR).
GN INSR.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hysticognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359245; PubMed=2768234;
RA Shier P., Watt V.M.;
RT "Primary structure of a putative receptor for a ligand of the insulin
RT family."
RL J. Biol. Chem. 264:14605-14608(1989).
CC -1- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN
CC AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE
CC -1- SUBUNIT: PROBABLE Tetramer of 2 alpha and 2 beta chains linked by
CC DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF
CC THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE
CC DOMAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: J05047; AAA37044.1; -
DR PIR: A36502; A36502.
DR HSSP: P06213; IIRK.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002011; Receptor_Tyr_kin_II.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00014; FNTPETIT.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00261; FU; 1.
DR SMART: SM00219; TYRKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

```

```

DR PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 26
FT CHAIN 27 746
FT CHAIN 747 1300
FT DOMAIN 747 921
FT TRANSMEM 922 943
FT DOMAIN 944 1300
FT DOMAIN 483 601
FT DOMAIN 813 910
FT DOMAIN 979 1254
FT NP_BIND 985 993
FT BINDING 1013 1013
FT ACT_SITE 1115 1115
FT MOD_RES 1145 1145
FT DISULFID 657 864
FT CARBOHYD 47 47
FT CARBOHYD 311 311
FT CARBOHYD 411 411
FT CARBOHYD 411 411
FT CARBOHYD 492 492
FT CARBOHYD 528 528
FT CARBOHYD 616 616
FT CARBOHYD 634 634
FT CARBOHYD 756 756
FT CARBOHYD 885 885
FT CARBOHYD 898 898
SQ SEQUENCE 1300 AA; 144517 MW; 3344DF4B6A57B24F CRC64;

```

```

alignment_scores:
Quality: 67.50 Length: 79
Ratio: 1.731 Gaps: 4
Percent Similarity: 49.367 Percent Identity: 29.114

```

alignment block:

US-09-823-101-2/rev x IRR_CAVPO ..

Align seg 1/1 to: IRR_CAVPO from: 1 to: 1300

```

195 GACACCTCCCTGATCTGTACCCACCACTTCTGCTCCACCTGCGCCA 146
    :::::||| ||||| |||||:::|:::|:::|:::|:::|:::|:::|
92 GluSerIleuArgAspIleuProAsnIleuAlaValIleArgIylAlaH 108
145 CTTCC.....TACTGCCTTGCTCTTCTCCAG.....G 120
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
108 sleuPheIeuGLYTYAlaIleuValIlePheGluMetProHISleuArgA 125
119 GTTGGATCTCCCTTCCTCGGACCTGTAAAGCATGTATT..... 79
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
125 spValIglyIeuProAlaIleuGLYAlaValIleuHISgluSerValArgVal 141
78 .....TGG...CTATG 71
142 GluIyAsnIgluIleuGLYleuGLYshISleuSerThIleAspTrpIglyIeu 158
70 AGAGCTGACTCTTATCTCTCTACCTGCTGACAC 34
158 uGlnProHnProSerThAsnTyrlIleValIglYasn 170

```

seq_name: SwissProt_39:SMD3_DROME

```

seq_documentation_block:
ID SMD3_DROME STANDARD; PRT; 151 AA.
AC O4437;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SMALL NUCLEAR RIBONUCLEOPROTEIN SM D3 (SNRP CORE PROTEIN D3) (SM-
D3).

```



```
CC RNA PROCESSING EVENTS). REGULATE ITS OWN EXPRESSION AT THE LEVEL
CC OF RNA PROCESSING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; SPECKLED SUBNUCLEAR COMPARTMENT.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: THREE MRNAs ARE PRODUCED DURING DEVELOPMENT.
CC THE SMALLEST OF THESE (3.5 KB RNAs) IS THE MAJORITY SPECIES DURING
CC PRECELLULAR BLASTODERM DEVELOPMENT AFTER WHICH ITS LEVELS DROP
CC RAPIDLY, BUT PERSISTS AS A MINORITY SPECIES THROUGHOUT THE REST OF
CC THE LIFE OF THE ORGANISM. THE LARGER TWO TRANSCRIPTS (4.4 AND 5.2
CC KB RNAs) FIRST APPEAR AROUND CELLULAR BLASTODERM AND LEVELS
CC INCREASE SUBSTANTIALLY DURING NEXT FEW HOURS AND ARE THE
CC PREPREDOMINANT RNA SPECIES THROUGHOUT THE REMAINDER OF THE LIFE OF
CC THE ORGANISM.
CC -1- DOMAIN: RS DOMAIN DIRECTS LOCALIZATION OF PROTEINS TO THE SPECKLED
CC SUBNUCLEAR COMPARTMENT AND THE PURPOSE OF THIS LOCALIZATION IS TO
CC ALLOW COLOCALIZATION AND CO-CONCENTRATION OF COMPONENTS OF THE
CC SPLICING AND SPLICING REGULATORY MACHINERY TO PERMIT RELATIVELY
CC HIGH RATES AND/OR EFFICIENCIES OF REACTION AND INTERACTION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way used
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; X06589; CAA29812.1; -.
DR EMBL; X06589; CAA29813.1; -.
DR PIR; S06028; S06028.
DR FlyBase; FBgn0003638; su(w[la]).
DR InterPro; IPR000061; Surp. 2.
DR Pfam; PF01805; Surp. 2.
KW Transcription regulation; RNA-binding; mRNA splicing; Repeat;
KW Nuclear protein; Alternative splicing.
FT DOMAIN 47 50 POLY-SRR.
FT DOMAIN 51 177 DRY CEERRL MOTIF.
FT DOMAIN 313 602 2 X REPEATS OF THE SURP MOTIF.
FT REPEAT 313 355 SURP MOTIF 1.
FT REPEAT 562 602 SURP MOTIF 2.
FT DOMAIN 528 533 POLY-GLU.
FT DOMAIN 640 643 POLY-ALA.
FT DOMAIN 826 829 POLY-ASP.
FT DOMAIN 829 964 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN 1035 1038 POLY-ARG.
FT VARSPLIC 100 177 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 1042 AA; 114928 MW; E564006EE163C3B3 CRC64;

Alignment_scores:
Quality: 67.00 Length: 99
Ratio: 1.426 Gaps: 3
Percent Similarity: 47.475 Percent Identity: 29.293

alignment_block:
US-09-823-101-2 x SUWA_DROME ..

Align seg 1/1 to: SUWA_DROME from: 1 to: 1042

6 TCACACATGCGCTGACTTGTCAGTGGGGGTGCCAAGAGGATAGAGAGA 55
:::||||| ||||| ||| ||| ||||| ||||| ::
384 TAAProSerAlaaspHisserIugluValAlaIaglylArgArGas 400
:::||||| ::||| |
56 TTAGAGTCAGCGCTGCATA..... 74
: ::|||::::|||
400 nProasnInuValIleThrValProThrIleIysTyrlsProserA 417
:::||||| ::||| ||||| ::||| |||
75 .....GCCAAATACTAATGCTTACAGTCCCGAGGAAGGAGAGATCCA... 116
::: ||| ||||| ::||| |||
417 IaasncyAlatylTrhglnleuIeserlyIleIyselylValProleu 433
117 .....AACCTGGAGAGACA.. 134
```

```

434 GlnAlaValLeuInclInuAspGluSerSerAspProGlyAsnSerGlnH1 450
135 .....GGCAGTAGAAGTGGGCGAGTGGGAGGACGAGAAAGCTTGGGTGA 177
      |||::: ||| ::::: |||:::
450 sserGlyGlyThrAlaSerProAlaLeuSerCysArgSerGluGlyHis 467
178 CAACATCAGGAGGAGGCTGTCTGACCTTTTCTTGTGAGGAATTCCTTAGC 224
      ::::: |||::: ||| ::::: |||
467 snserGlnGlyGlyGluPheThrProValLeuLeuGlnInTyAsnGly 482

seq_name: SwissProt_39:FBRL_SCHPO

seq_documentation_block:
ID   FBRL_SCHPO          STANDARD:          PRT;          305 AA.
AC   P35551;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   FIBRILLARIN.
GN   FIB OR SPBC2D10.10C.
OS   Schizosaccharomyces pombe (fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=93361822; PubMed=8493104;
RA   Girard J.-P., Fellu J., Calzergues-Ferrer M., Lapeyre B.;
RT   "Study of multiple fibrillar mRNAs reveals that 3' end formation in
RN   Schizosaccharomyces pombe is sensitive to cold shock.";
RN   Nucleic Acids Res. 21:1881-1887(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RL   Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
RA   Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: FIBRILLARIN IS A COMPONENT OF A NUCLEOLAR SMALL NUCLEAR
CC   RIBONUCLEOPROTEIN PARTICLE THOUGHT TO PARTICIPATE IN THE FIRST
CC   STEP IN PROCESSING PERIOSOMAL RNA.
CC   -1- SUBCELLULAR LOCATION: NUCLEAR; FIBRILLAR REGION OF THE NUCLEOLUS.
CC   -1- PTM: BY HOMOLOGY TO OTHER FIBRILLARINS, SOME OR ALL OF THE
CC   N-TERMINAL ARGININES ARE N,N-DIMETHYLATED (DMA).
CC   -1- SIMILARITY: BELONGS TO THE FIBRILLARIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on ways
CC   the European Bioinformatics Institute. There are no restrictions on ways
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
CC   EMBL; X69930; CAA49550.1; -.
DR   EMBL; AL031788; CAA21168.1; -.
DR   PIR; S33690; S33690.
DR   InterPro: IPR000692; FibrillarIn.
DR   Pfam; PF00269; FibrillarIn_1.
DR   PRINTS; PR00052; FIBRILLARIN.
DR   ProDom; PD004637; FibrillarIn_1.
DR   PROSITE; PS00566; FIBRILLARIN; 1.
KW   Nuclear protein; Methylation; Ribonucleoprotein; rRNA processing;
KW   RNA-binding.
FT   DOMAIN            6          70          GLY/DMA-RICH.
SQ   SEQUENCE 305 AA; 32040 MW; 5AE68B37EC3331 CRC64;

Alignment_scores:
Quality: 66.50          Length: 54
Ratio: 2.463           Gaps: 2
Percent Similarity: 50.000 Percent Identity: 38.889

```

alignment_block:

US-09-823-101-2 x FBRL_SCHPO

Align seg 1/1 to: FBRL_SCHPO from: 1 to: 305

```
30 GGGGGTGTCCAGAGAGATAGAGATAGAGCTCAGGCTCATAGCCAA 79
||||| :|||
19 GtlyGlyPheAnsnglyGlyArg..... 25
80 ATACTATGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAGAGA 129
||||| :|||
26 .....GtlyGlyPheGlyGlyArgGlyGlyAlaArgGlyGlyGly 40
130 GACAAGCGCAGTAGAAGTGGCGAGCGAGCAAGAGCTTGCGTGACA 179
||| :|||
40 rGtlyGlyAlaArgGlyGlyArg...GtlyGlyArgGlyGlyAlaArgGly 55
180 GATCAGCGAGG 191
:::|||||
56 GtlyArgGlyGly 59
```

seq_name: Swissprot_39:LRPL_HSVLF

seq_documentation_block:

```
ID LRPL_HSVLF STANDARD; PRT; 340 AA.
AC P17588;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE LATENCY-RELATED PROTEIN 1.
OS Herpes simplex virus (type 1 / strain F).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10304;
```

```
RM SEQUENCE FROM N.A.
RP MEDLINE=89085598; PubMed=2535901;
RA Wehlsler S.L., Nesburn A.B., Zwaagstra J., Ghiasi H.;
RT "Sequence of the latency-related gene of herpes simplex virus type
1."
RL Virology 168:168-172(1989).
```

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

```
DR EMBL: J04323; AAA45799.1; -.
KW PIR: A33337; WMBELI.
FT Repeat.
FT DOMAIN 27 75 2 x 17 AA REPEATS.
FT REPEAT 27 43 1.
FT REPEAT 59 75 2.
SQ SEQUENCE 340 AA; 35604 MM; 2977FA8F87E5451 CRC64;
```

alignment_scores:

Quality: 66.50 Length: 44
Ratio: 3.167 Gaps: 1
Percent Similarity: 47.727 Percent Identity: 40.909

alignment_block:

US-09-823-101-2 x LRPL_HSVLF

Align seg 1/1 to: LRPL_HSVLF from: 1 to: 340

```
30 GGGGGTGTCCAGAGAGATAGAGATAGAGCTCAGGCTCATAGCCAA 79
||||| :|||
299 GtlyGlyArgGlyGlyArgGly..... 307
```

```
80 ATACTATGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAGAGA 129
||||| :|||
308 .....SerArgGlyArgGlyGlyArgGlyArgGlyGly 319
130 GACAAGCGCAGTAGAAGTGGCGAGCGAGCAAGAGCTTGCGTGACA 161
||| :|||
319 rGtlyGlyAlaArgArgGlyArgGlyGlyGly 329
```

seq_name: Swissprot_39:ROG_MOUSE

seq_documentation_block:

```
ID ROG_MOUSE STANDARD; PRT; 388 AA.
AC O35479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN G (HNRP G).
GN RBMXPI OR RBMXRT OR HNRP OR HNRPg.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RM SEQUENCE FROM N.A.
RP MEDLINE=98119027; PubMed=9457683;
RA Delbridge M.L., Ma K., Subbarao M.N., Cooke H.J., Bhasin S.,
RA Graves J.A.M.;
RT "Evolution of mammalian HNRPg and its relationship with the putative
azooasperma factor RBM."
RL Mamm. Genome 9:168-170(1998).
CC -1- FUNCTION: BINDS RNA.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.
CC -1- PTM: O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RBM).
```

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

```
DR EMBL: AF031568; AAB86639.1; -.
DR MGD: MG1:1343045; Rbmxt.
DR InterPro: IPR00504; RBM.
DR Pfam: PF00076; Rbm; 1.
DR SMART: SM00360; RBM; 1.
DR PROSITE: PS00102; RBM; 1.
DR PROSITE: PS00030; RBM_RNP_1; 1.
DR KW Nuclear protein; RNA-binding; Ribonucleoprotein; Glycoprotein.
FT DOMAIN 8 86
SQ SEQUENCE 388 AA; 42233 MM; F02F2DCB8C87F14 CRC64;
```

alignment_scores:

Quality: 66.50 Length: 68
Ratio: 2.375 Gaps: 2
Percent Similarity: 41.176 Percent Identity: 32.353

alignment_block:

US-09-823-101-2 x ROG_MOUSE

Align seg 1/1 to: ROG_MOUSE from: 1 to: 388

```
42 GGAGGTAGAGAGATAGAGCTCAGGCTCATAGCCAAATACTAT..... 86
||||| :|||
318 GtlyGlySerArgAspSerArgSerArgSerArgSerArgSerArgSerArg 334
87 .....GGCTACAGTCCGAGCA..... 104
334 rGtlyArgAspArgValGtlyArgGlnGlnArgGlyLeuProProSerMetG 351
```

105AGGGA 110
 351 LuArgGLYTYRProProARgAspSerTYRserSerSerArgly 367
 111 GATCAAAACCTCGAAGACAGCAGTGAAGTGGCGAGTGGAGG 160
 368 AlaProARgGLYGLYArgGLYGLYSerARgSerARgGLYGLY 384
 161 CAGG 164
 384 YArg 385

seq_name: SwissProt_39: ROG_HUMAN

seq_documentation_block:
 ID ROG_HUMAN STANDARD; PRT; 391 AA.
 AC P38159;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN G (HNRP G) (GLYCOPROTEIN P43).
 GN RBMX1 OR HNRP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast;
 RX MEDLINE=94021365; PubMed=7692398;
 RA Soulard M., della Valle V., Stomi M., Pinol-Roma S., Codogno P., Bauvy C., Belli M., Lacroix J.-C., Monod G., Dreyfuss G., Larsen C.-J.,
 RT "hnRNP G: sequence and characterization of a glycosylated RNA-binding protein.";
 RT Nucleic Acids Res. 21:4210-4217(1993).
 RL [2]
 RA Revisions.
 RP Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS RNA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.
 CC -1- PTM: O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z23064; CAA80599.1; -
 CC DR PIR: S43202; S43202.
 CC DR HSP: P09651; IUP1.
 CC MIM: 300199; -
 CC DR InterPro: IPR000504; RRM.
 CC DR Pfam: PF00076; rrm_1.
 CC DR SMART: SM00360; RRM; 1.
 CC DR PROSITE: PS50102; RRM; 1.
 CC DR PROSITE: PS00030; RRM_RNP_1; 1.
 CC KW Nuclear protein; RNA-binding; Ribonucleoprotein; Glycoprotein.
 CC FT DOMAIN 8 86
 CC SEQUENCE 391 AA; 42403 MW; 904FE911FBD3626 CRC64;

alignment_scores: 66.50 Length: 68
 Ratio: 2.375 Gaps: 2
 Percent Similarity: 41.176 Percent Identity: 32.353

alignment_block:
 US-09-823-101-2 x ROG_HUMAN ..

Align seg 1/1 to: ROG_HUMAN from: 1 to: 391

42 GGAGTAGAGAGATAGAGAGTCAAGCTCATAGCCAAATACAT..... 86
 321 GLYLYSerARgAspSerTYRserSerSerArgSerpleuTYRser 337
 87GCCTTACAGTCCGAGCA..... 104
 337 rGLYArgAspArgValGLYArgGLNGLYArgLeuProPserMetG 354
 105AGGGA 110
 354 LuArgGLYTYRProProARgAspSerTYRserSerSerArgly 370
 111 GATCAAAACCTCGAAGACAGCAGTGAAGTGGCGAGTGGAGG 160
 371 AlaProARgGLYGLYArgGLYGLYSerARgSerARgGLYGLY 387
 161 CAGG 164
 387 YArg 388

seq_name: SwissProt_39: RS2_ARATH

seq_documentation_block:
 ID RS2_ARATH STANDARD; PRT; 285 AA.
 AC P49688; O22936;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 40S RIBOSOMAL PROTEIN S2.
 GN RPS2 OR AT2G41840 OR T11A07.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V., Buehl C.R., Ketchum K.A., Lee J.J., Konning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman J.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.,
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.";
 RT Nature 402:761-768(1999).
 RL [2]
 RP SEQUENCE OF 119-285 FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Raynal M., Grellier F., Laudie M., Meyer Y., Cooke R., Delzeny M.; Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AC002339; AAC02764.1; -
 CC DR EMBL: Z17622; CAA79017.1; -

DR	EMBL	X66899	CAA47350.1	-
DR	EMBL	X72990	CAA51489.1	-
DR	EMBL	X72991	CAA51489.1	JOINED
DR	EMBL	X72992	CAA51489.1	JOINED
DR	EMBL	X72993	CAA51489.1	JOINED
DR	EMBL	X72994	CAA51489.1	JOINED
DR	EMBL	X72995	CAA51489.1	JOINED
DR	EMBL	X72996	CAA51489.1	JOINED
DR	EMBL	X72997	CAA51489.1	JOINED
DR	EMBL	X72998	CAA51489.1	JOINED
DR	EMBL	X72999	CAA51489.1	JOINED
DR	EMBL	X73000	CAA51489.1	JOINED
DR	EMBL	X73001	CAA51489.1	JOINED
DR	EMBL	X73002	CAA51489.1	JOINED
DR	EMBL	X73003	CAA51489.1	JOINED
DR	EMBL	X73004	CAA51489.1	JOINED
DR	EMBL	X08806	CAA70044.1	ALT_INT.

SEQUENCE FROM N.A.
TISSUE=Testis;

RA Atkinson M.R., Townsend-Nicholson A., Nicholl J.K., Sutherland G.R.,
 RA Schofield P.R.;
 CC Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
 CC IS MEDIATED BY G PROTEINS WHICH INHIBITS ADENYLATE CYCLASE.
 CC POSSIBLE ROLE IN REPRODUCTION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L20463; AAA16365.1; -;
 DR EMBL; L22607; AAA35949.1; -;
 DR EMBL; X76981; CAA54288.1; -;
 DR EMBL; L77730; AAB02790.1; -;
 DR EMBL; L77729; AAB02790.1; JOINED.
 DR PIR; S38511; S38511.
 DR HSSP; P29274; 1MMH.
 DR GCRDB; GCR_0746; -;
 DR GCRDB; GCR_0791; -;
 DR GCRDB; GCR_1878; -;
 DR GCRDB; GCR_2061; -;
 DR MIM; 600445; -;
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PRO0237; GPCRHHODPSN.
 DR PRINTS; PRO0424; ADENOSINER.
 DR PRINTS; PRO0555; ADENOSINER3.
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR_F1_1; 1.
 DR PROSITE; PS50262; G_PROTEIN_RECEPTOR_F1_2; 1.
 DR KMW; G-protein coupled receptor; Transmembrane; Glycoprotein;
 KM Lipoprotein; Palmitate.
 FT DOMAIN 1 14 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 15 37 1 (POTENTIAL).
 FT DOMAIN 38 48 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 49 72 2 (POTENTIAL).
 FT DOMAIN 73 84 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 85 106 3 (POTENTIAL).
 FT DOMAIN 107 126 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 127 148 4 (POTENTIAL).
 FT DOMAIN 149 177 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 178 198 5 (POTENTIAL).
 FT DOMAIN 199 231 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 232 255 6 (POTENTIAL).
 FT DOMAIN 256 261 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 262 284 7 (POTENTIAL).
 FT DOMAIN 285 318 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 83 166 BY SIMILARITY.
 FT LIPID 303 303 PALMITATE (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 12 12 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 7 7 A -> T (IN REF. 1).
 FT CONFLICT 248 248 I -> L (IN REF. 3).
 SQ SEQUENCE 318 AA; 36184 MW; 690E67986130FC28 CRC64;

alignment_scores:
 Quality: 65.50 Length: 54
 Ratio: 2.113 Gaps: 2
 Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:
 US-09-823-101-2/rev x AA3R_HUMAN ..

Align seg 1/1 to: AA3R_HUMAN from: 1 to: 318

seq_name: SwissProt_39:FBRL_YEAST
 seq_documentation_block:
 ID FBRL_YEAST STANDARD; PRT; 327 AA.
 AC P15646; P89890;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FIBRILLARIN (NUCLEOLAR PROTEIN 1).
 GN NOP1 OR YDL014W OR D2870.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BJ2168;
 RX MEDLINE=90130477; PubMed=2298745;
 RA Henriquez R., Blobel G., Aris J.P.;
 RT "Isolation and sequencing of NOP1. A yeast gene encoding a nucleolar
 RT protein homologous to a human autoimmune antigen.";
 RL J. Biol. Chem. 265:2209-2215(1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC MEDLINE=90076121; PubMed=2686980;
 RX Schlimmang T., Tollervey D., Kern H., Frank R., Hurt E.C.;
 RA "A yeast nucleolar protein related to mammalian fibrillarin is
 RT associated with small nucleolar RNA and is essential for viability.";
 RL EMBO J. 8:4015-4024(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Andre B., Vissers S., Urrestarazu L.;
 RT submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP FUNCTION.
 RX MEDLINE=91160511; PubMed=1825809;
 RA Tollervey D., Lelton H., Carmo-Fonseca M., Hurt E.C.;
 RT "The small nucleolar RNP protein NOP1 (fibrillarin) is required for
 RT pre-rRNA processing in yeast.";
 RL EMBO J. 10:573-583(1991).
 CC -1- FUNCTION: FIBRILLARIN IS A COMPONENT OF A NUCLEOLAR SMALL NUCLEAR
 CC RIBONUCLEOPROTEIN PARTICLE THOUGHT TO PARTICIPATE IN THE FIRST
 CC STEP IN PROCESSING PRE-RIBOSOMAL RNA.
 CC -1- SUBUNIT: ASSOCIATES WITH NOP56 AND NOP58.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; FIBRILLARIN REGION OF THE NUCLEOLUS.
 CC -1- PTM: BY HOMOLOGY TO OTHER FIBRILLARINS, SOME OR ALL OF THE
 CC N-TERMINAL ARGININES ARE N-DIMETHYLATED (DMA).
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLARIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: J05230; AAA34816.1; -
DR EMBL: 248432; CA88345.1; -
DR EMBL: 274061; CA98571.1; -
DR EMBL: 274062; CA98572.1; -
DR PIR: A35038; A35038.
DR PIR: S25421; S25421.
DR SGD: S0002172; NOP1.
DR InterPro: IPR000692; Fibrillarin.
DR Pfam: PF01269; Fibrillarin; 1.
DR PRINTS: PR00052; FIBRILLARIN.
DR PRODOM: PD004637; Fibrillarin; 1.
DR PROSITE: PS00566; Fibrillarin; 1.
KW Nuclear protein; Methylation; Ribonucleoprotein; rRNA processing;
KW RNA-binding.
FT DOMAIN 6 83 GLY/DNA-RICH.
FT DOMAIN 22 42 RNA-BINDING RGG-BOX (BY SIMILARITY).
FT DOMAIN 58 75 RNA-BINDING RGG-BOX (BY SIMILARITY).
FT DOMAIN 281 313 HELICAL (POTENTIAL).
SQ SEQUENCE 327 AA; 34465 MW; 56A8B958A7B066E CRC64;

```

```

alignment_scores:
Quality: 65.50 Length: 50
Ratio: 2.339 Gaps: 3
Percent Similarity: 56.000 Percent Identity: 40.000

```

```

alignment_block:
US-09-823-101-2 x FBRL_YEAST ..

```

```

Align seg 1/1 to: FBRL_YEAST from: 1 to: 327

```

```

30 GGGGGTCCAGAGAGTAGAGAGATACAGTCAGCTCATAGCCAA 79
||||| ::||| ||| |||:::
41 Glylyserarglyglyalalarglyglyserargly..... 53
80 ATACTATGCTTACAGTCCGAGAGAGGAGATCAACCTGGAAGAA 129
::||| ||||| ::|||
54 .glyphely.....glyargglyglyserarglyglyalaa 66
130 GACAGGACAGTAGAGAGTGGCGA.....GTGGAGGAGGAGAAAGT 170
|| |||||::||| ||||| ::|||
66 rgglyglyserarglyglyargglyglyalaaaglyglyalaaagly 82

```

```

seq_name: SwissProt_39:FUS_BOVIN

```

```

seq_documentation_block:

```

```

ID FUS_BOVIN STANDARD; PRT; 512 AA.

```

```

AC 028009;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-BINDING PROTEIN FUS (PIGEPEN PROTEIN).
GN FUS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=orta;
RX MEDLINE=96175600; PubMed=8631501;
RA Allegro M.C.; Allegro M.A.;
RT "A nuclear protein regulated during the transition from active to
RT quiescent phenotype in cultured endothelial cells.";
RL Dev. Biol. 174:288-297(1996).

```

```

RN [2]
RX CARBOHYDRATE BINDING DOMAIN.
RX MEDLINE=20160719; PubMed=10694442;
RA Allegro M.C.;
RT "A C-terminal carbohydrate-binding domain in the endothelial cell
RT regulatory protein, pigpen: new function for an EWS family member.";

```

```

RL Exp. Cell Res. 255:270-277(2000).
CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND
CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
CC STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED
CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR, EXHIBITS DIFFUSE STAINING
CC THROUGHOUT (EXCLUDING NUCLEOLI), TOGETHER WITH A SMALL NUMBER OF
CC INTENSELY STAINED FOCAL POINTS, OR GRANULES, AND PUNCTATE STAINING
CC ALONG THE NUCLEAR ENVELOPE.
CC -1- DOMAIN: THE C-TERMINAL DOMAIN BINDS CARBOHYDRATES.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL: U26024; AAC13543.1; -
CC DR HSP: P09651; 1HA1.
CC DR InterPro: IPR000504; RRM.
CC DR InterPro: IPR001876; znf-RanBP.
CC Pfam: PF00076; rrm; 1.
CC DR Pfam: PF00641; znf-RanBP; 1.
CC DR SMART: SM00360; RRM; 1.
CC DR SMART: SM00547; znf-RBZ; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 164 GLN/GLY/SER/TYR-RICH.
FT DOMAIN 165 253 GLY-RICH.
FT DOMAIN 271 357 RNA-BINDING (RRM).
FT DOMAIN 357 512 ARG/GLY-RICH.
FT ZN_FING 414 433 C4-TYPE (POTENTIAL).
SQ SEQUENCE 512 AA; 52240 MW; 3652329C04F1386 CRC64;

```

```

alignment_scores:
Quality: 65.50 Length: 76
Ratio: 1.770 Gaps: 5
Percent Similarity: 48.684 Percent Identity: 32.895

```

```

alignment_block:
US-09-823-101-2 x FUS_BOVIN ..

```

```

Align seg 1/1 to: FUS_BOVIN from: 1 to: 512

```

```

27 AGTGGGGTTCAGAGAGTAGAGAGATACAGTCAGCTCATAGC 76
||||| |||||::||| ||| |||:::
197 Serglyglytyrglyglyglnglnaspargly..... 208
77 CAATACTATGCTTACAGTCCGAGAGAGGAGATCCA..... 116
209 .....glyargglyargglyglyglytyrtyra 219
117 .....AACCTGGAGAGACAGACAGAGAGAGT 146
219 snargserglyglytyrgluProargglyargglyglytyrtyrtyr 235
147 GGGGAGTGGAGGACAGAAAGTTGGGTGACATCAGAGAGAGTCT 196
||||| ||||| ||| ::|||::||| |||
236 glyarg...glyglymetglygly.....Seraspargglygly..... 247
197 GACCTTTTCTTGAGAGAAATTTTAGGC 224
248 .....Pheasnlyspheglygly 253

```

THIS PAGE BLANK (USPTO)

us-09-823-101-2.n2p.rspt

Species	SP_rodent:Q9Z0G8	SP_human:Q9Y4G0	SP_invertebrate:Q9YVZ2	SP_invertebrate:Q9Y1D2	SP_fungi:Q9C2Q9
<i>rattus norvegicus</i> (rat)	-	67.00	124.24	24.79	485
<i>homo sapiens</i> (human)	-	67.00	118.88	24.36	864
<i>drosophila melanog</i>	+	67.00	118.53	24.23	1032
<i>drosophila melanog</i>	+	67.00	118.47	24.23	1041
<i>neurospora crassa</i>	+	66.50	130.37	29.15	188

seq_name:	sp_rodent:09WVC9				
seq_documentation_block:					
ID	09WVC9	PRELIMINARY;	PRT;	873	AA.
AC	09WVC9;				
DT	01-NOV-1999	(TREMBlrel. 12, Created)			
DT	01-NOV-1999	(TREMBlrel. 12, Last sequence update)			
DT	01-JUN-2001	(TREMBlrel. 17, Last annotation update)			
DE	HUNTINGTON YEAST PARTNER C.				
GN	2610317D23RK1 OR HYPC.				
OS	Mus musculus (mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCHI_taxid:	10090;				
XX					

RP SEQUENCE FROM N.A.

RA Bedford M.T., Das R., Reed R., Leder P.;

RT "EBP11, a mammalian ortholog of the essential yeast splicing factor
RT PRP40.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF135440: AAD39464.1: -

DR MGD; MGI:1925583; 2610317D23R1k.

```

DR Pfam; PF00397; MW: 2.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00441; EF: 4.
DR SMART; SM00456; MW: 2.
DR PROSITE; PS01159; MW_DOMAIN_1; 1.
DR PROSITE; PS50020; MW_DOMAIN_2; 2.
DR SEQUENCE 873 AA; 99697 MW; E3FCFDBC3725A5PF CRC64;

alignment_scores:
    Quality: 78.00
    Ratio: 2.889
    Percent Similarity: 49.091
    Percent Identity: 32.727

alignment_block:
US-09-823-101-2/rev x Q9WVC9 ..

Align seg 1/1 to: Q9WVC9 from: 1 to: 873

190 CCTCCCTGATCTGTGCACCCACCTTCTCGCCGACACGCGCCACTTCT 141
    ||||| ||| ||| ::||| ||| ||| ::
22 ProPrometSergInaPrgProProAlaIleProPrometProGlyI11 38
    ||||| ::||| ||| ::||| ||| |||
140 ACTGCCTGTCTTCTTCCACGAGGTTTGATCTCCCTCCGCGGACTGTA 91
    ||||| ::||| ||| ::||| ||| |||
38 eleuProPrometIleuProPrometGlyAlaProPro..... 51
90 AGCCATGATATTGGCTATGAGAGCCCTGACTTATCTCTCTACCTCT 41
    ||||| ::|||
52 .....leuThrlInIleProGly 57
40 TGGACACCCCGCAGTC 26
    ::||| ||| ::
58 MetValProPromet 62

seq_name: sp_invertebrate:Q9UAY0

seq_documentation_block:
ID Q9UAY0 PRELIMINARY; PRT; 471 AA.
AC Q9UAY0;
QT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

```

DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE W03G1.5 PROTEIN.
GN W03G1.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Pauley A., Scheet P., Harper M.;
RT "The sequence of C. elegans cosmid W03G1."
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF125964; AAD14753.1; -; BDF30B59A64A9B5 CRC64;
SQ SEQUENCE 471 AA; 50885 MW; BDF30B59A64A9B5 CRC64;

alignment_scores:
Quality: 75.50 Length: 49
Ratio: 2.517 Gaps: 2
Percent Similarity: 61.224 Percent Identity: 42.857

alignment_block:
US-09-823-101-2 x Q9UAY0 ..
Align seg 1/1 to: Q9UAY0 from: 1 to: 471

39 CAAGGAGGTAGAGAGTAAAGTCAAGCTCTCATAGCCAAATACATATG 88
AC ||||| :|||: :|||: |||
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DE 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE 01-NOV-1996 (TREMBlrel. 01, last annotation update)
DE B2126.F3.142.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RL Smith D.R.;
RT Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: U00017; AAA17220.1; -;
SQ SEQUENCE 298 AA; 30858 MW; B3573B073E2139C3 CRC64;

alignment_scores:
Quality: 75.00 Length: 55
Ratio: 2.143 Gaps: 1
Percent Similarity: 63.636 Percent Identity: 34.545

alignment_block:
US-09-823-101-2/rev x Q49806 ..
Align seg 1/1 to: Q49806 from: 1 to: 298

185 CTGATCTGTACCCACACTTTCCTCCGCTCCGACCTGACCTTACTGC 136
AC ||||| :|||: :|||: |||
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DE 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE SIMILAR TO NUCLEOLIN.
GN C27H5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```


RT elegans." ;
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL NZ;
 RA Pauley A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL NZ;
 RA Waterston R.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U14635; AAC4657.1; -.
 DR InterPro: IPR000504; RRM.
 DR InterPro: IPR001876; Znf-RanBP.
 DR Pfam: PF00076; rtm; 1.
 DR Pfam: PF00641; zf-RanBP; 1.
 DR PROSITE: PS0102; RRM; 1.
 DR SMART: SM00360; RRM; 1.
 DR SMART: SM00547; Znf_RB2; 1.
 SQ SEQUENCE 545 AA; 56896 MW; 36D771CDA39A0CAE CRC64;

alignment_scores:
 Quality: 75.00 Length: 65
 Ratio: 2.419 Gaps: 3
 Percent Similarity: 47.692 Percent Identity: 38.462

alignment_block:
 US-09-823-101-2 * Q18265 ..

Align seg 1/1 to: Q18265 from: 1 to: 545

```

27 AGTGGGGGTGTCACAGAGTAGAGATAGAGTCAGGCTCTCATAGC 76
   |||||||  ::|||
123 SeeglyglInsergly.....Se 130
   |||||||  |||
77 CAAATCTAGTGTAGCTTACCCGAGAGGAGATCCAAAC..... 119
   :      |||||||  |||
130 rasprotyrgly..GlnserarglyglyargglyPhcgly 146
120 .....CCTGAAGAAGACAGAGGCACTAGACT 146
   |||  |||  |||||||
146 lyserarglyglyglytyrtyrarglyglyserargly 162
147 GGGCGAGTGGAGCGCAGAAAGTGGTGACAGACAGAGG 191
   |||  |||||||  |||
163 glytyrarglyglytyrtyrarglyglytyrarglygly 177

```

seq_name: sp_Invertebrate:Q9V4J6

seq_documentation_block:

```

ID Q9V4J6 PRELIMINARY: PRT: 1307 AA.
AC Q9V4J6:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG11101 PROTEIN.
GN CG11101.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dublin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Pauleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195(2000).
DR EMBL: AE003841; AAF59274.2; -
DR HSSP: P35555; IEMN.
DR FlyBase: FBgn003157; CG11101.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF00008; EGF; 2.
DR PRINTS: PRO1217; PRICHEXTENS.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR PROSITE: PS01187; EGF_CA; 1.
DR KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1307 AA; 139890 MW; 42827878C1B7510 CRC64;

```

alignment_scores:
 Quality: 75.00 Length: 85
 Ratio: 1.744 Gaps: 5
 Percent Similarity: 50.588 Percent Identity: 29.412

alignment_block:
 US-09-823-101-2/rev x Q9V4J6 ..

Align seg 1/1 to: Q9V4J6 from: 1 to: 1307

```

229 CACTTCCCTAAGAAATTCCTCAAGAAAG.....GTCAACA 192
   ::|||  ::|||
235 AsnleuprolleaspyrgruAlaGlnLysProGlyThrselwlrgrp 251
191 CCTCCCTGATCTGTCAACCAACCTTCTGCTCCCACTCGCCACTTC 142
   |||||  |||||
251 opfopro.....TyrProheasparpgrogluAtrpoph 264
141 TACTGCGTGTCTTCTTCAGAGTTTGATCTCCCTCTCGCGAGCTGT 92
   :      |||  |||
264 lthr.....GlyvalproleprogluGlnlle 273

```

```

91 AAGCCATAGTATTGGCTATGAGAGCCGTACTTATCTCTCCACTCC 42
   |||
   :|:|:|
274 ValPro.....ValargProGlyPheValSerAsnArgProPr 286
41 TTGG.....ACACCCCACTGCACAGTCAGCCAGCTGGG 7
   |||
286 OTIPYrArgSnlYsProArgProProIleAlaThrSerValGlyGlyA 303
6 AGAGG 2
   :|:|:|
303 snArg 304

```

seq_name: sp_human:094945

```

seq_documentation_block:
ID 094945 PRELIMINARY; PRT; 526 AA.
AC 094945;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0867 PROTEIN.
GN KIAA0867.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=9156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364 (1998).
DR EMBL: AB020674; BAA74890.1;
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF00010; HLH; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00353; HLH; 1.
SQ SEQUENCE 526 AA; 56822 MW; 0C3ADB0852823B60 CRC64;

```

alignment_scores:

Quality:	73.00	Length:	51
Ratio:	2.920	Gaps:	3
Percent Similarity:	49.020	Percent Identity:	43.137

alignment_block:

US-09-823-101-2/rev x 094945 ..

Align seg 1/1 to: 094945 from: 1 to: 526

```

190 CCTCCCTGATCTGTACCCCACTTTCCTGCTCC...ACTGCCCACT 144
   |||
   :|:|:|
27 ProProLeuSerValProGlnProPheLeuProValPheThrMetProLe 43
143 TCCTACTGCTTGTCTTCTCCAGGGTTGGATCTCCCTTCCTCGGGACT 94
   |||
   :|:|:|
43 uLeuSerPro.....SerProAlaProProProI 53
GN
93 GTAAGCATAAGTATTGGCTATGAGAGCCGTACTTATCTCTCCACT 44
   :|:|:|
53 leSerPro.....ValLeuProLeuValPro 61
43 CCT 41
   |||
62 Pro 62

```

seq_name: sp_human:09HAP2

```

seq_documentation_block:
ID 09HAP2 PRELIMINARY; PRT; 919 AA.
AC 09HAP2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MONDA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20528627; PubMed=11073985;
RA Billin A.N., Eilers A.L., Coulter K.L., Logan J.S., Ayer D.E.;
RT "MonDA, a Novel Basic Helix-Loop-Helix-Leucine Zipper Transcriptional
RT Activator That Constitutes a Positive Branch of a Max-Like Network.";
RL Mol. Cell. Biol. 20:8845-8854 (2000).
DR EMBL: AF312918; AAG34121.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 919 AA; 10113 MW; 47375FC8B6CAAC2 CRC64;

```

alignment_scores:

Quality:	73.00	Length:	51
Ratio:	2.920	Gaps:	3
Percent Similarity:	49.020	Percent Identity:	43.137

alignment_block:

US-09-823-101-2/rev x 09HAP2 ..

Align seg 1/1 to: 09HAP2 from: 1 to: 919

```

190 CCTCCCTGATCTGTACCCCACTTTCCTGCTCC...ACTGCCCACT 144
   |||
   :|:|:|
420 ProProLeuSerValProGlnProPheLeuProValPheThrMetProLe 436
143 TCCTACTGCTTGTCTTCTCCAGGGTTGGATCTCCCTTCCTCGGGACT 94
   |||
   :|:|:|
436 uLeuSerPro.....SerProAlaProProProI 446
GN
93 GTAAGCATAAGTATTGGCTATGAGAGCCGTACTTATCTCTCCACT 44
   :|:|:|
446 leSerPro.....ValLeuProLeuValPro 454
43 CCT 41
   |||
455 Pro 455

```

seq_name: sp_invertebrate:09VU42

```

seq_documentation_block:
ID 09VU42 PRELIMINARY; PRT; 1406 AA.
AC 09VU42;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG10600 PROTEIN.
GN CG10600.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Basu P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borckova K., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasse K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert R., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "the genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL: AE003660; AAF53715.2; -.
DR Flybase: FBgn0032717; CG10600.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 1406 AA; 153642 MW; 928070CF036CF4D3 CRC64;

alignment_scores:
Quality: 73.00 Length: 70
Ratio: 2.281 Gaps: 2
Percent Similarity: 45.714 Percent Identity: 30.000

alignment_block:

US-09-823-101-2/rev x Q9VJ42 ..

Align seg 1/1 to: Q9VJ42 from: 1 to: 1406

226 TTGCCTAGAGATTCTCTCACAAGAAAAAGTCAGACACCTCCCTGATCTGT 177
:::|||||:::|||||
1245 ILeProLysSerPheLeu.....HisProProValProI 1256
:::|||||
176 CACCCAACTTTCTGCGCTCCACAGTGGCCACTGTACAGCGCTGTGCTTC 127
:::|||||
1256 eValAlaAlaProAlaProProProSerInIleValLysValL 1273
:::|||||
1273 euProGlySerLeuThrProLeuThrArg..... 1282
:::|||||
76 GCTATGAGAGCCTGACTTATCTCTCTGACCTCTTGAGACACCCCACT 27
:::|||||
1283MetSerValLLeuProProLeuProProLe 1293
:::|||||
26 CACCAAGTCA 17
|||
1293 uProProSer 1296

seq_name: sp_plant:Q9AWI3
seq_documentation_block:
ID Q9AWI3 PRELIMINARY; PRT: 234 AA.
AC Q9AWI3;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE P0489A05.17 PROTEIN.
GN P0489A05.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriophytidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV; NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0489A05.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003105; BAB32994.1; -.
SQ SEQUENCE 234 AA; 24551 MW; ABCDB8DB100D26E CRC64;

alignment_scores:
Quality: 72.00 Length: 49
Ratio: 2.323 Gaps: 1
Percent Similarity: 63.265 Percent Identity: 34.694

alignment_block:

US-09-823-101-2 x Q9AWI3 ..

Align seg 1/1 to: Q9AWI3 from: 1 to: 234

45 GGTAGAGAGATAGATGAGTGCAGCTCTCATAGCCAAATACCTAGCTTACA 94
|||||:::|||||
155 GLYArgArgCysSerArgProAlaValAlaIleThrArgSerAlaIleAl 171
:::|||||
95 GTCCCGAGAGAGGAGATCCAAACCTCGAGAGACAGCAGTACGAGA 144
:::|||||
171 agInArgGlyGluGluGluArgGluGluGly.....GlyGluArgL 186
:::|||||
145 GTGGCGAGCTGGAGGAGCAGAAAGTTGGTGCACAGATCAGGAGGG 191
:::|||||
186 yGclYwctGluGluArgLysGlyTTPArgSerAsnGluArgGly 201
:::|||||
seq_name: sp_plant:Q41848
seq_documentation_block:
ID Q41848 PRELIMINARY; PRT: 301 AA.
AC Q41848;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE PROLIN RICH PROTEIN.
GN PRP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W64A;
RA MEDLINE=92361259; PubMed=1498600;
RA Jose-Estanyol M., Ruiz-Avila L., Puigdomenech P.;
RT "A maize embryo-specific gene encodes a proline-rich and hydrophobic
protein.";
RL Plant Cell 4:413-423(1992).
DR EMBL: X60432; CAA42959.1; -.


```

33 GGTGTCCAGAGAGTAGAGATAGACTGCTCTCATGACCAATA 82
   ||| :|||:| ||| ||| :||| :||| :||| :||| :|||
146 GlyGlyArgGlyGlyTyrArgAspGlyGlyArg.....GlyGlyTyr 159
83 CTAATGGCTTACAGTCCCGAGAGGAGATCCCAACCTCGAAGAGAC 132
   | ||| :|||:| ||| ||| ||| :||| :||| :||| :|||
159 TgLyGlyTyrArgAspGlyGlyArgGlyGlyGlyTyrArgA 176
133 AAGCAGTAGAGAGTGG..... 149
   :|||:| |||:| ||| :||| :||| :||| :||| :|||
176 spGlyGlyArgGlyGlyTyrArgAspGlyGlyTyrArgAspGlyGly 192
150 ...CGAGTGGAGCAGCAAGAGTGGTGACATCAGAGAGG 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 TyrArgAspGlyGlyArgGlyGlyTyrGlyGlyGlyArgGlyGly 207
seq_name: sp_Invertebrate:Q9GV11

```

```

seq_documentation_block:
ID Q9GV11 PRELIMINARY; PRT; 491 AA.
AC Q9GV11;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE P110-RELATED PROTEIN POP110 (FRAGMENT).
GN POP110.
OS Ephydatia fluviatilis.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haplosclerida; Spongillidae; Ephydatia.
OX NCBI_TaxID=31330;
RN [1]
RP SEQUENCE FROM N.A.
RA Mochizuki K., Fujisawa T.;
RT "P110-related gene in sponge.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
CC -1- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR EMBL; AB047384; BAB13309.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD. 1.
DR Pfam; PF00271; Helicase_C. 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR KW ARP-binding; Helicase.
FT NON_TER 1
SQ SEQUENCE 491 AA; 54545 MM; C77CACB33B642FE3 CRC64;

```

```

alignment_scores:
Quality: 70.50 Length: 60
Ratio: 2.518 Gaps: 1
Percent Similarity: 46.667 Percent Identity: 36.667

```

alignment_block:

US-09-823-101-2 x Q9GV11 ..

Align seg 1/1 to: Q9GV11 from: 1 to: 491

```

27 AGTGGGGTGTCCAGAGAGTAGAGATTAAGATCAGGCTCTCATAC 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 SerGlyGlyGlyPheGlyGlyArgAspTyrArg..... 451
77 CAATACTATGTGGCTTACAGTCCCGAGAGGAGATCCAAACCTCGAA 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
452 .....GlnSerArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 463
127 GAAGACAGCAGTAGAAGTGGCGAGTGGAGCGAGCAAGAGTTGGTGG 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 spRtyGlyTyrSerArgGlyGlyAlaGlyGlyGlyArgGlyGlyMetSer 479
177 ACAGATCAGAGAGGTGTCTGACCTTTTTC 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

480 GlyGlyGlyGlyGlySerValAspTyrTrp 489
seq_name: sp_Invertebrate:O62213

```

```

seq_documentation_block:
ID O62213 PRELIMINARY; PRT; 208 AA.
AC O62213;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F33A8.3 PROTEIN.
GN F33A8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitida; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Faveilo A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
DR EMBL; Z81525; CAB04257.1; -.
DR HSSP; P41016; IC90.
DR InterPro; IPR002059; Cold_shock.
DR Pfam; PF00313; CSD; 1.
DR PRINTS; PR00050; COLDSHOCK.
DR PRODOM; PD000621; Cold_shock; 1.
DR SMART; SM00357; CSP; 1.
DR PROSITE; PS00352; COLD_SHOCK; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 208 AA; 21349 MM; 2A6C81AAE3D85FC8 CRC64;

```

```

alignment_scores:
Quality: 70.00 Length: 77
Ratio: 1.892 Gaps: 4
Percent Similarity: 48.052 Percent Identity: 35.065

```

alignment_block:

US-09-823-101-2 x O62213 ..

Align seg 1/1 to: O62213 from: 1 to: 208

```

30 GGGGGTGTCCAGAGAGTAGAG.....AGAGATACAGTAGGC 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
93 GlyProValGlnGlySerItyrAlaAlaAspArgAspAlaGluAsnAl 109
68 TTCATACAGCAATACTATAGGCTTACAGTCCCGAGAGAG..... 107
109 A.....AlaArgGlyArgGlyGlyArgGlyArgGlyArgGlyGly 124
108 ..... 110
124 rgGlyGlyLleArgHisAspSerGlySerArgAspAlaGluGluGlyGly 140
111 GATCCAAACCTCGAAGAAAGACAGCAGTACAGTGGCGAGTGGAGG 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

141 AlaProAArgGlyGlyArgGlyGlySerArgArgGlyGly... Gly1 156
 161 CAGGAAGGTTGGTGACAGATCAGGAGG 191
 |||||
 156 yArgGlyGlyGlyArgGlyThrAsnSerGlyGly 166

127 GAAGCAAGCGAGTAGAGTGGCGAGTGGAGCGAGGAAGGT 170
 |||||
 171 spArgTyrglySerArgSerGlyArgAspGlyGlyArgGly 185

seq_name: sp_Plant:Q42412

seq_documentation_block:

ID Q42412 PRELIMINARY; PRT; 209 AA.

AC Q42412;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE RNA-BINDING PROTEIN R2-1.

OS Nicotiana sylvestris (Wood tobacco).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:

OC Asteridae: euasterids I; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxId=4096;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96397973; Pubmed=8804857;

RA Hanano S., Sugita M., Sugitara M.;

RT "Structure and expression of the tobacco nuclear gene encoding RNA-

binding protein R2-1: the existence of an intron in the 3'-

untranslated region."

RL DNA Res. 3:65-71(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-YOUNG LEAF;

RA Sugitara M.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-YOUNG LEAF;

RA Hanano S., Sugita M., Sugitara M.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.

DR EMBL; D83696; BAA12064.1; -.

DR EMBL; D28861; BAA06012.1; -.

DR Mendel; 15272; Nicsy:2406;15272.

DR InterPro: IPR000504; RRM.

DR InterPro: IPR001878; Znf_CCHC.

DR Pfam; PF00076; rrm; 1.

DR Pfam; PF00098; zf-CCHC; 1.

DR PRINTS; PR00939; C2HCZNFINGER.

DR SMART; SM00360; RRM; 1.

DR SMART; SM00343; ZNF_C2HC; 1.

DR PROSITE; PS50102; RRM; 1.

DR PROSITE; PS00030; RRM_RNP_1; 1.

DR RNA-binding; Zinc-finger.

SQ SEQUENCE 209 AA; 22457 MW; 72BFF5EA263569F5 CRC64;

alignment_scores:

Quality: 70.00 Length: 48

Ratio: 2.692 Gaps: 1

Percent Similarity: 54.167 Percent Identity: 41.667

alignment_block:

US-09-823-101-2 x Q42412 ..

Align seg 1/1 to: Q42412 from: 1 to: 209

27 AGTGGGGCTGCCAGAGGTAGAGATGAAGCTCAGGCTCATAGC 76
 |||||
 144 SerGlyGlyGlyArgGlyArgTyrglyGlyGlyGly..... 157
 77 CAATACTACTAGCTTACAGTCCGAGAGGAGGATCCAAACCCGTGAA 126
 |||||
 158GlySerArgSerSerGlyTyrglyProAspArgAsnGlyA 171

THIS PAGE BLANK (USPTO)


```

FEATURES
  source
    location/Qualifiers
      1..829
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
      34..42
        /note="variant of CAAT-box"
      64..69
        /note="Goldberg-Hogness box"
      151..472
        /note="exon 1 (472 is 1st base in codon)"
      151..796
        /note="putative primary transcript"
      join(262..472,603..697)
        /note="precursor"
      /codon_start=1
      /protein_id="CAA25005.1"
      /db_xref="GI:31649"
      /db_xref="SWISS-PROT:P01350"
      /translation="MORLCYVYLIFALALAFSEASMKPRSQPDALPGTGANDRLLEI
        PWLDOOGASHHRRLQGLGQGPPLVADPSKKGWPLLEEEBAVGMNDRGRASDEN
        262 318
        join(436..472,603..667)
        /note="Protein sequence is in conflict with the conceptual
        translation"
      /codon_start=1
      /product="G34 (big gastrin)"
      /protein_id="CAA25006.1"
      /db_xref="GI:1335072"
      /db_xref="SWISS-PROT:P01350"

FEATURES
  source
    location/Qualifiers
      1..829
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
      264..437
        /product="human gastrin signal peptide"
      join(438..474,605..699)
        /product="human gastrin mature peptide"
      700..831
        /product="human gastrin mature peptide"
      3'UTR

FEATURES
  source
    location/Qualifiers
      1..829
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
      199 a
        250 c
        213 g
        167 t

BASE COUNT
  199 a
  250 c
  213 g
  167 t

ORIGIN
  199 a
  250 c
  213 g
  167 t

Query Match
  Best Local Similarity 45.3%; Score 47.6; DB 6; Length 829;
  Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

40 gatgcagcagcagatgtgtatgtatgtctgtctcttgcactagatcagcgcctctctg 99
  |||||||
261 GATGCAGCAGCAGCATGTGTATGTATGTCTGTCTCTGACTTTCGCACATCGCCCTTCTCG 319
  |||||||

Oy 100 aacct 105
  |||||

Db 320 AAGCTT 325

RESULT 4
LOCUS HSGAS1 829 bp DNA PRI 10-FEB-1999
DEFINITION Human gastrin gene.
ACCESSION X00183
VERSION X00183.1 GI:31648
KEYWORDS gastrin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
  Kato,K., Hayashizaki,Y., Takahashi,Y., Himeno,S. and Matsubara,K.
  Molecular cloning of the human gastrin gene
  Nucleic Acids Res. 11 (23), 8197-8203 (1983)
  8419471
FEATURES
  source
    location/Qualifiers
      1..829
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
      34..42
        /note="variant of CAAT-box"
      64..69
        /note="Goldberg-Hogness box"
      151..472
        /note="exon 1 (472 is 1st base in codon)"
      151..796
        /note="putative primary transcript"
      join(262..472,603..697)
        /note="precursor"
      /codon_start=1
      /protein_id="CAA25005.1"
      /db_xref="GI:31649"
      /db_xref="SWISS-PROT:P01350"
      /translation="MORLCYVYLIFALALAFSEASMKPRSQPDALPGTGANDRLLEI
        PWLDOOGASHHRRLQGLGQGPPLVADPSKKGWPLLEEEBAVGMNDRGRASDEN
        262 318
        join(436..472,603..667)
        /note="Protein sequence is in conflict with the conceptual
        translation"
      /codon_start=1
      /product="G34 (big gastrin)"
      /protein_id="CAA25006.1"
      /db_xref="GI:1335072"
      /db_xref="SWISS-PROT:P01350"

sig_peptide
  CDS
    sig_peptide
    CDS

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 61; Conservative	92.48; Pred. No. 1.2e-05;	45.38; Score 47.6; DB 9;	Length 829;	
		Mismatches 4; Indels 1; Gaps 1;		
QY 40	gatgcagcgaagatgtgtagtgcctctcttcagtagatcagccgcctctctg	99		
Db 261	GATGCGAGCATATGTGTGTATGTCTCATCTTTCACCT-GGCTCTGGCCGCTTCTCTG	319		
QY 100	aagctt 105			
Db 320	AAGCTT 325			
RESULT 5				
HUNGAST2	1217 bp	DNA		
LOCUS	Human gastrin gene, complete cds.		PRI	08-AUG-1995
DEFINITION	K01254 J00147			
ACCESSION	K01254.1 GI:182987			
VERSION	Alu repeat; gastrin; repeat region.			
KEYWORDS	2 of 2			
SEGMENT	Human fetal liver genomic DNA (Charon 4A library of Lawn et al.),			
SOURCE	and gastrinoma, cDNA to mRNA (see comment).			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 601 to 815; 946 to 1136)			
	Kato, K., Himeno, S., Takahashi, Y., Wakabayashi, T., Tarui, S. and			
	Matsudaira, K.			
TITLE	Molecular cloning of human gastrin precursor cDNA			
JOURNAL	Gene 26 (1), 53-57 (1983)			
MEDLINE	84159488			
REFERENCE	2 (bases 600 to 815; 946 to 1139)			
AUTHORS	Boel, E., Vuust, J., Norris, F., Norris, K., Wind, A., Rehfeld, J. F. and			
	Merckel, K. A.			
TITLE	Molecular cloning of human gastrin cDNA: evidence for evolution of			
JOURNAL	gastrin by gene duplication			
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2866-2869 (1983)			
REFERENCE	3 (bases 471 to 1151)			
AUTHORS	Wiborg, O., Berglund, L., Boel, E., Norris, F., Norris, K.,			
	Rehfeld, J. F., Merckel, K. A. and Vuust, J.			
TITLE	Structure of a human gastrin gene			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 81 (4), 1067-1069 (1984)			
MEDLINE	84144842			
REFERENCE	4 (bases 1 to 1217)			
AUTHORS	Ito, R., Sato, K., Helmer, T., Jay, G. and Agrawal, K.			
TITLE	Structural analysis of the gene encoding human gastrin: the large			
JOURNAL	intron contains an Alu sequence			
	Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4662-4666 (1984)			

THIS PAGE BLANK (USPTO)

Search for 09823101

Your SELECT statement is:

s ssg and cancer

Items File

3 5: Biosis Previews(R)_1969-2003/Mar W2
4 34: SciSearch(R) Cited Ref Sci_1990-2003/Mar W2
4 71: ELSEVIER BIOBASE_1994-2003/Mar W2
30 73: EMBASE_1974-2003/Mar W2
1 98: General Sci Abs/Full-Text_1984-2003/Feb
1 144: Pascal_1973-2003/Mar W1
5 149: TGG Health&Wellness DB(SM)_1976-2003/Feb W4
2 155: MEDLINE(R)_1966-2003/Mar W2
1 159: Cancerlit_1975-2002/Oct
1 162: CAB Health_1983-2003/Jan
2 266: FEDRIP_2003/Jan

SYSTEM:OS - DIALOG OneSearch

File 73:EMBASE 1974-2003/Mar W2

(c) 2003 Elsevier Science B.V.

*File 73: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.

File 159:Cancerlit 1975-2002/Oct

(c) format only 2002 Dialog Corporation

*File 159: Cancerlit ceases updating with immediate effect.

Please see HELP NEWS.

File 155:MEDLINE(R) 1966-2003/Mar W2

(c) format only 2003 The Dialog Corp.

Set Items Description

S1 388 SSG

S2 395880 STOMACH OR GASTRIC

S3 1 S1 AND S2

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 08:48:53 ; Search time 2206.96 Seconds
(without alignments)
784.882 Million cell updates/sec

Title: US-09-823-101-7

Perfect score: 105
Sequence: 1 tgcagtggtgcagacatcc.....ggccgcctctctgaagctt 105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_da:*
3: gb_da:*
4: gb_da:*
5: gb_da:*
6: gb_da:*
7: gb_da:*
8: gb_da:*
9: gb_da:*
10: gb_da:*
11: gb_da:*
12: gb_da:*
13: gb_da:*
14: gb_da:*
15: gb_da:*
16: gb_da:*
17: gb_da:*
18: gb_da:*
19: gb_da:*
20: gb_da:*
21: gb_da:*
22: gb_da:*
23: gb_da:*
24: gb_da:*
25: gb_da:*
26: gb_da:*
27: gb_da:*
28: gb_da:*
29: gb_da:*
30: gb_da:*
31: gb_da:*
32: gb_da:*
33: gb_da:*
34: gb_da:*
35: gb_da:*
36: gb_da:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47.6	45.3	405	6	E01795	E01795 mRNA sequen
2	47.6	45.3	413	9	HSRASP	V00511 Human mRNA
3	47.6	45.3	829	6	E01796	E01796 DNA sequenc
4	47.6	45.3	829	9	HSRASP	X00183 Human gastr
5	47.6	45.3	1217	9	HMGAST2	K01254 Human gastr
6	47.6	45.3	7739	9	HMGASTA	M15958 Human gastr
7	47.6	45.3	451	4	ECPPGAST	Y09440 Equus cabal
8	47.6	45.3	457	4	ECPPGAST	X15682 Feline mRNA
9	47.6	45.3	447	4	ECPPGAST	M16581 Bovine gastr
10	47.6	45.3	1066	4	BOVGAS	M31657 Bovine gastr
11	47.6	45.3	32	4	SCGAST	V01303 Porcine mes
12	47.6	45.3	30.5	4	OU92801	U92801 Ovis aries
13	47.6	45.3	30.2	4	H0M2B5SH12	AF086156 Homo sapi
14	47.6	45.3	685	9	HSRASP	AL110142 Homo sapi
15	47.6	45.3	1751	9	BC004912	BC004912 Homo sapi
16	47.6	45.3	2857	9	AK023487	AK023487 Homo sapi
17	47.6	45.3	3964	9	AB018271	AB018271 Homo sapi
18	47.6	45.3	67033	9	AL137008	AL137008 Human DNA
19	47.6	45.3	28.6	9	AL138894	AL138894 Human DNA
20	47.6	45.3	32719	2	AC010733	AC010733 Homo sapi
21	47.6	45.3	130000	9	AC087259	AC087259 Homo sapi
22	47.6	45.3	28.4	9	AC073504	AC073504 Homo sapi
23	47.6	45.3	28.4	8	AP002483	AP002483 Oryza sat
24	47.6	45.3	175565	8	AP003311	AP003311 Oryza sat
25	47.6	45.3	28.0	9	AL391361	AL391361 Human DNA
26	47.6	45.3	161079	2	AC083842	AC083842 Homo sapi
27	47.6	45.3	28.0	2	AC073167	AC073167 Homo sapi
28	47.6	45.3	182105	2	AC036149	AC036149 Homo sapi
29	47.6	45.3	28.0	2	AC015984	AC015984 Homo sapi
30	47.6	45.3	17250	2	AC027264	AC027264 Homo sapi
31	47.6	45.3	190748	9	AC010969	AC010969 Homo sapi
32	47.6	45.3	208558	2	AL392043	AL392043 Homo sapi
33	47.6	45.3	151	11	G65603	G65603 sy2795 YAC
34	47.6	45.3	27.4	153	G65864	G65864 sy698 MISCe
35	47.6	45.3	27.4	460	MMGASTGNI	X94758 M.musculus
36	47.6	45.3	1055	10	MMGASTGNI	X94760 M.musculus
37	47.6	45.3	1267	10	MMU58136	U58136 Mus musculu
38	47.6	45.3	5154	10	MMU34293	U34293 Mus musculu
39	47.6	45.3	137682	9	AC023274	AC023274 Homo sapi
40	47.6	45.3	172811	2	AC080086	AC080086 Homo sapi
41	47.6	45.3	177421	2	AC032013	AC032013 Mus muscu
42	47.6	45.3	177769	9	AC006328	AC006328 Homo sapi
43	47.6	45.3	273521	2	AL590968	AL590968 Mus muscu
44	47.6	45.3	13521	1	AE003904	AE003904 Xylella f
45	47.6	45.3	35240	2	AC040987	AC040987 Homo sapi

ALIGNMENTS

RESULT	1	LOCUS	E01795	405 bp	RNA	PAT	29-SEP-1997
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
COMMENT							
PD							

PF 24-JUN-1987 JP 1987156745
PI SATO TAKAKI, HIMENO SEIICHI, HAYASHIZAKI YOSHITIDE, PI
KATOU KIKUYA,
PI MATSUBARA KENICHI
PC C12N15/00.C07H21/04.C12N1/16.C12P21/02.(C12P21/02.C12R1:865);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue=Stomach;
CC *source: cell-type=Antrum mucosal tissue;
CC *source: clone=PHG 53;
FH Key Location/Qualifiers
FH 5'UTR 1..4
FT sig_peptide 5..61
FT /product='human gastrin signal peptide' FT
FT mat_peptide 62..172
FT /product='human gastrin precursor' FT
FT misc_signal 173..178
FT /note='region deleted by tripsin-like FT
FT protease' /note='region deleted by tripsin-like FT
FT mat_peptide 179..307
FT /product='human gastrin'
FT misc_signal 224..229
FT /note='region deleted by tripsin-like FT
FT protease' /note='region deleted by tripsin-like FT
FT misc_signal 284..289
FT /note='region deleted by tripsin-like FT
FT 3'UTR 308..404
FT polyA_signal 378..381
FT polyA_signal 385..390.
location/Qualifiers
1..405
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 99 a 123 c 110 g 73 t
ORIGIN

Query Match 45.3%; Score 47.6; DB 6; Length 405;
Best Local Similarity 92.4%; Pred. No. 1.4e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 40 gatcgacgacgagtgtgtatgtctgtcctcttcacatagatcagccgctctctg 99
|||||
Db 4 GATCGACGACGACTATGTGTATGTGCTGACTTTCGACT-GGCTCTGCCGCTTCTCTG 62
|||||

QY 100 aagctt 105
|||||

Db 63 AAGCTT 68

RESULT 2
LOCUS HSGAST 413 bp mRNA PRI 21-NOV-1994
DEFINITION Human mRNA encoding pregastrin (a regulatory hormone of gastric acid secretion and growth of the gastrointestinal mucosa).
ACCESSION V00511
VERSION V00511.1 GI:31654
KEYWORDS complementary DNA; gastrin; signal peptide.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 413)
Boel,E., Vuust,J., Norris,F., Norris,K., Wind,A., Rehfeld,J.F. and Marcker,K.A.
Molecular cloning of human gastrin cDNA: evidence for evolution of gastrin by gene duplication
Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2866-2869 (1983)
JOURNAL
MEDLINE 83221503

COMMENT Data kindly reviewed (21-JUN-1983) by Esper Boel.
FEATURES
source Location/Qualifiers
1..413
/organism='Homo sapiens'
/db_xref='taxon:9606'
mRNA
1..413
/product='pregastrin'
CDS
9..314
/codon_start=1
/product='pregastrin'
/protein_id='CAA23769.1'
/db_xref='GI:31653'
/translation='MORLCVYLIFALALAAFSASWKPDSQDPAPLGTGANRLLEL
PMLEOQGPASHHRQLGPGPPHLVADPSKKOGPMLEEFEEAYGWMDFRRSAEDEN'
sig_peptide 9..65
polyA_site 413
BASE COUNT 102 a 127 c 110 g 74 t
ORIGIN

Query Match 45.3%; Score 47.6; DB 9; Length 413;
Best Local Similarity 92.4%; Pred. No. 1.4e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 40 gatcgacgacgagtgtgtatgtctgtcctcttcacatagatcagccgctctctg 99
|||||
Db 8 GATCGACGACGACTATGTGTATGTGCTGACTTTCGACT-GGCTCTGCCGCTTCTCTG 66
|||||

QY 100 aagctt 105
|||||

Db 67 AAGCTT 72

RESULT 3
LOCUS E01796 829 bp DNA PAT 29-SEP-1997
DEFINITION DNA sequence coding for human gastrin.
ACCESSION E01796
VERSION E01796.1 GI:2170048
KEYWORDS JP 1989002579-A/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 829)
Sato,T., Himeno,S., Hayashizaki,Y., Katou,K. and Matsubara,K.
RECOMBINANT DNA, MICROORGANISM CONTAINING SAID DNA AND PRODUCTION OF POLYPEPTIDE THEREWITH
Patent: JP 1989002579-A 2 06-JAN-1989;
JOURNAL AJINOMOTO CO INC
OS Homo sapiens (Human)
PN JP 1989002579-A/2
PD 06-JAN-1989
PI SATO TAKAKI, HIMENO SEIICHI, HAYASHIZAKI YOSHITIDE, PI
KATOU KIKUYA,
PI MATSUBARA KENICHI
PC C12N15/00.C07H21/04.C12N1/16.C12P21/02.(C12P21/02.C12R1:865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue=Stomach;
CC *source: cell-type=Antrum mucosal tissue;
CC *source: clone=plamdaHG 11;
FH Key Location/Qualifiers
FH 5'UTR 1..263
FT sig_peptide 34..42
FT misc_signal 64..69
FT /standard_name='GCCAATCT model sequence' FT
FT misc_signal 64..69
FT /standard_name='Gold berg-Hogness box' FT

gene /note="gastrin, (first expressed exon); G00-119-261"
/number=2
6289. .6499
/gene="GAS"
CDS join(6289. .6499,6630. .6724)
/note="gastrin"
/codon_start=1
/protein_id="AA52520.1"
/db_xref="GI:182991"
Intron /translation="MORLCYVILFALALAFSEASMKPSQOPDAPLGANDLEL
PMLRQGPASHRRQLGPPHLPADPSKKGFWLEEEBAATGMDFGRRSAEDEN
6500. .6629
/note="gastrin cds intron B"
exon 6630. .>6724
/note="gastrin"
repeat_region /note="Alu-9 repeat"
7005. .7349
/number=3
BASE COUNT 1964 a 2036 c 1844 g 1895 t
ORIGIN 1 bp upstream of EcoRI site; chromosome 17q.

Query Match 45.3%; Score 47.6; DB 9; Length 7739; *
Best Local Similarity 92.4%; Pred. No. 9.5e-06;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 40 gatcagcagcagatggtatgtatgtctcttgcactagatcagccgctctctg 99
|||||
Db 6288 GATCAGCGACTATGTGTATGTCTCTGACTTCTGACTTCTGACTTCTG 6346
|||||

OY 100 aagctt 105
|||||
Db 6347 AAGCTT 6352

RESULT 7
ECPPGAST
LOCUS ECPPGAST 451 bp mRNA MAM 19-AUG-1998
DEFINITION Equus caballus mRNA for preprogastrin.
ACCESSION Y09440
VERSION Y09440.1 GI:3445273
KEYWORDS gast gene; preprogastrin.
SOURCE horse.
ORGANISM Equus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE 1 (bases 1 to 451)
Johnsen, A.H., Sandin, A., Rourke, I.J., Bundgaard, J.R., Nilsson, G.
and Rehfeld, J.F.
Unique progastrin processing in equine G-cells suggests marginal
tyrosyl sulfotransferase activity
Eur. J. Biochem. 255 (2), 432-438 (1998)
JOURNAL MEDLINE 96380242
REFERENCE 2 (bases 1 to 451)
Rourke, I.J.
Direct Submission
Submitted (14-NOV-1996) I. J. Rourke, Department of Clinical
Biochemistry, National University Hospital, Blegdamsvej 9,
Copenhagen, DK-2100, DENMARK
JOURNAL MEDLINE
REFERENCE 3 (bases 1 to 451)
Revised by (3)
Rourke, I.J.
Direct Submission
Submitted (17-AUG-1998) I. J. Rourke, Department of Clinical
Biochemistry, National University Hospital, Blegdamsvej 9,
Copenhagen, DK-2100, DENMARK
JOURNAL MEDLINE
REMARK 3 (bases 1 to 451)
TITLE Direct Submission
AUTHORS Rourke, I.J.
COMMENT Submitted (17-AUG-1998) I. J. Rourke, Department of Clinical
Biochemistry, National University Hospital, Blegdamsvej 9,
Copenhagen, DK-2100, DENMARK
FEATURES
source location/Qualifiers
1. .451
/organism="Equus caballus"
/db_xref="taxon:9796"
/dev_stage="adult"
/tissue_type="antral mucosa"

gene 41. .364
/gene="gast"
41. .364
CDS /gene="gast"
/codon_start=1
/product="preprogastrin"
/protein_id="CAA70590.1"
/db_xref="GI:1684683"
mat_peptide /db_xref="SWISS-PROT:P55885"
HWPQDLRLGPAASHRRQLGLQSGPHLVADLSKKGFWLEEEBAATGMDFGRRSAE
GDQSP"
224. .328
/evidence="experimental"
/product="mature gastrin-34 coding sequence including
amidation substrate"
429. .434
polya_signal
BASE COUNT 102 a 142 c 127 g 80 t
ORIGIN

Query Match 39.6%; Score 41.6; DB 4; Length 451;
Best Local Similarity 85.3%; Pred. No. 0.0015;
Matches 58; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

OY 38 atgacgacgacgagtgtatgtatgtctcttgcactagatcagccgctctc 97
|||||
Db 38 AAGATGCGCGCAGCTGTGTATGTCTGATCTTGGCTTGGCTTGGCTTGG 96
|||||

OY 98 tgaagctt 105
|||||
Db 97 CGAAGCTT 104

RESULT 8
FSPPGAST
LOCUS FSPPGAST 457 bp mRNA MAM 19-MAR-1991
DEFINITION Feline mRNA for preprogastrin.
ACCESSION X16582
VERSION X16582.1 GI:1099
KEYWORDS gastrin; preprogastrin.
SOURCE wild cat.
ORGANISM Felis silvestris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE 1 (bases 1 to 457)
Kim, S.J.
Direct Submission
Submitted (22-FEP-1989) Kim S. J., Korea Advanced Institute of
Science and Technology, Department of Biological Science and
Engineering, P.O. Box 150, Chongryang, Seoul 131, Korea
JOURNAL MEDLINE 92127058
REFERENCE 2 (bases 1 to 457)
Kim, S.J., Uhm, K.N., Kang, Y.K. and Yoo, O.J.
Bovine and feline gastrin cDNA sequences and the amino acid and
nucleotide sequence homologies among mammalian species
DNA Seq. 1 (3), 181-187 (1991)
JOURNAL MEDLINE 92127058
FEATURES
source location/Qualifiers
1. .457
/organism="Felis silvestris"
/db_xref="taxon:9683"
/dev_stage="adult"
47. .103
47. .361
/codon_start=1
/product="preprogastrin"
/protein_id="CAA34599.1"
/db_xref="GI:1100"
/db_xref="SWISS-PROT:P01354"
/translation="MORLCYVILFALALAFSEASMKPSQODAPSGGANGLEP
HMLNRLGPASHRRQLGPPHLPADPSKKGFWLEEEBAATGMDFGRRSAEDCQD
HP"

Db 596 CGACCTT 603

|||||

RESULT 11

SCGAST 462 bp mRNA MAM 01-OCT-1996

LOCUS Porcine messenger RNA for preprogastrin.

DEFINITION V01303 J00651

ACCESSION V01303.1 GI:1846

VERSION V01303.1 GI:1846

KEYWORDS complementary DNA; gastrin.

SOURCE Sus scrofa

ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 462)

AUTHORS Yoo,O.-J., Powell,C.T. and Agarwal,K.L.

TITLE Molecular cloning and nucleotide sequence of full-length of CDNA coding for porcine gastrin

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 79 (4), 1049-1053 (1982)

FEATURES

Location/Qualifiers

1..462

/organism="Sus scrofa"

/db_xref="taxon:9823"

1..462

/note="messenger RNA"

62..376

/note="reading frame preprogastrin"

/codon_start=1

/protein_id="CAA24610.1"

/db_xref="GI:1847"

/translation="MQRCAVLIHLVLAALACASASMKPGFOLDASSGPGANGKEP
HELDRLGPASHRRRLGIQDPPLVADLAKKQGWMEEEEAAYGMDFGRSAEEDG
RP"

BASE COUNT 110 a 147 c 135 g 70 t

ORIGIN

Query Match 30.5%; Score 32; DB 4; Length 462;

Best Local Similarity 76.5%; Pred. No. 2.9;

Matches 52; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 38 atgagcagcagcagatgtgtatgtctgtccttgcactagatcagccgccttc 97

Db 59 AAGATGACAGCGCTGTCGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 117

QY 98 tgaagctt 105

Db 118 TGAAGCTT 125

RESULT 12

OAU92801 746 bp DNA MAM 16-JUL-1999

LOCUS Ovis aries preprogastrin (GAS) gene, exons 2, 3 and complete cds.

DEFINITION O92801

ACCESSION O92801.1 GI:1917024

VERSION O92801.1 GI:1917024

KEYWORDS

SOURCE sheep.

ORGANISM Ovis aries

REFERENCE 1 (bases 1 to 746)

AUTHORS Moore,C., Jie,R., Shulkes,A. and Baldwin,G.S.

TITLE Molecular cloning and sequence of the ovine gastrin gene

JOURNAL DNA Seq. 8 (1-2), 39-44 (1997)

PUBMED 9522119

REFERENCE 2 (bases 1 to 746)

AUTHORS Moore,C., Ren,J., Shulkes,A. and Baldwin,G.S.

|||||

Direct Submission

JOURNAL Submitted (10-MAR-1997) Surgery, ASRMC, Studley Road, Heidelberg, Vic. 3084, Australia

FEATURES

Location/Qualifiers

1..746

/organism="Ovis aries"

/db_xref="taxon:9940"

/cell_type="leukocytes"

<1..>746

/gene="GAS"

<1..99

/gene="GAS"

/number=1

/join(<100..315,462..>647)

/gene="GAS"

/product="preprogastrin"

100..315

/gene="GAS"

/number=2

/join(105..315,462..565)

/gene="GAS"

/note="precursor of the hormone gastrin"

/codon_start=1

/product="preprogastrin"

/protein_id="AAB51307.1"

/db_xref="GI:1917025"

/translation="MQRCAVLIHLVLAALACASASMKPGFOLDASSGPGANGKEP
HELDRLGPASHRRRLGIQDPPLVADLAKKQGWMEEEEAAYGMDFGRSAEEDG
HP"

316..461

/gene="GAS"

/number=2

462..>647

/gene="GAS"

/number=3

BASE COUNT 150 a 241 c 221 g 134 t

ORIGIN

Query Match 30.5%; Score 32; DB 4; Length 746;

Best Local Similarity 76.5%; Pred. No. 2.7;

Matches 52; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 38 atgagcagcagcagatgtgtatgtctgtccttgcactagatcagccgccttc 97

Db 102 AAGATGACAGCGCTGTCGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 160

QY 98 tgaagctt 105

Db 161 CGAAGCTT 168

RESULT 13

HUMZB55H12 632 bp mRNA PRI 29-AUG-1998

LOCUS Homo sapiens full length insert CDNA clone ZB55H12.

DEFINITION AF086156

ACCESSION AF086156.1 GI:3483501

VERSION AF086156.1 GI:3483501

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 632)

AUTHORS Moessner,J., Tan,F., Marra,M., Kucaba,T., Tandel,M., Martin,J.,
Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behmer,K., Hillier,L.,
Wilson,R. and Waterston,R.

TITLE Full Clone Sequencing of the longest Available Member from Each
Unigene Cluster

JOURNAL unpublished

REFERENCE 2 (bases 1 to 632)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (24-AUG-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT SUBMITTED BY: Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES
 source Location/Qualifiers
 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Soares_fetal_lung_NbHL19W"
 /clone_image="IMAGE:307559"

BASE COUNT 183 a 117 c 123 g 209 t
 ORIGIN

Query Match 28.8%; Score 30.2; DB 9; Length 632;
 Best Local Similarity 69.5%; Pred. No. 11;
 Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 Oy 15 acatccctgcagctctcgtgacgatgtagcagcagatgtgtatgtctgtcttt 73
 ||||||| ||||||| ||| ||| ||||| ||||| ||||| ||||| |||||
 Db 281 ACTTCCTGTGAGTTCCTGTGCAGTACTGCMAAGACTGCGTGGTGTGTTGCTTT 339

RESULT 14
 LOCUS HSM800786 685 bp mRNA PRI 18-FEB-2000
 DEFINITION Homo sapiens mRNA: cDNA DKFZp564D0264 (from clone DKFZp564D0264).
 ACCESSION AL110142
 VERSION AL110142.1 GI:5817037
 KEYWORDS human.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 685)
 Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann,S.

TITLE Direct Submission
 JOURNAL Submitted (15-AUG-1999) MIPS, Am Klopferstr. 18a, D-82152 Martinsried, GERMANY

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; sequenced by Olegen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp564D0264) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
 source Location/Qualifiers
 1..685

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp564D0264"
 /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host X1-2blue; sites NotI + SalI"
 /dev_stage="fetal"
 /tissue_type="brain"
 648..653
 polyA_signal
 polyA_site
 666
 BASE COUNT 239 a 107 c 112 g 227 t
 ORIGIN

Query Match 28.8%; Score 30.2; DB 9; Length 685;
 Best Local Similarity 69.5%; Pred. No. 11;
 Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 15 acatccctgcagctctcgtgacgatgtagcagcagatgtgtatgtctgtcttt 73
 ||||||| ||||||| ||| ||| ||||| ||||| ||||| ||||| |||||
 Db 170 ACTTCCTGTGAGTTCCTGTGCAGTACTGCMAAGACTGCGTGGTGTGTTGCTTT 112

RESULT 15
 LOCUS BC004912/c 1751 bp mRNA PRI 12-JUL-2001
 DEFINITION Homo sapiens, clone IMAGE:3534745, mRNA, partial cds.
 ACCESSION BC004912
 VERSION BC004912.1 GI:13436208
 KEYWORDS human.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1751)

REFERENCE 1 (bases 1 to 1751)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadnan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 11 Row: a Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES
 source Location/Qualifiers
 1..1751

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3534745"
 /tissue_type="lung, small cell carcinoma"
 /clone_lib="NIH-MGC.7"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 <1..788
 /codon_start=3
 /product="Unknown (protein for IMAGE:3534745)"
 /protein_id="AAH04912.1"
 /db_xref="GI:13436209"
 /translation="HEVTRQVAKCKAKRFQVEIGIDNRYRFQDSSQQLRVILRSTV

CDS
 1..788
 /product="Unknown (protein for IMAGE:3534745)"
 /protein_id="AAH04912.1"
 /db_xref="GI:13436209"
 /translation="HEVTRQVAKCKAKRFQVEIGIDNRYRFQDSSQQLRVILRSTV

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:42:46 ; Search time 2644.98 Seconds
(without alignments)
438.772 Million cell updates/sec

Title: US-09-823-101-6
Perfect score: 108
Sequence: 1 acctatttgatattgagga.....aaaaaaaaaaaaaacgcgcg 108

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estlpl:*
6: em_estlba:*
7: em_estro:*
8: em_estov:*
9: em_hlcc:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	87	80.6	219 10	AM022375 df37902.y
2	87	80.6	392 10	AM022874 df45d04.y
3	87	80.6	447 10	AM055633 EST367703
4	85.4	79.1	409 11	N68114 y254a12.s1
5	85.4	79.1	591 10	AM069257 cr42c12.x
6	84.6	78.3	452 10	AM022830 df42c12.x
7	84.4	78.1	184 10	AM059779 tt86c08.x
8	83.8	77.6	433 10	AM0594372 hg58d07.x
9	83.8	77.6	543 10	AM0594372 hg58d07.x
10	83.6	77.4	275 11	AM0594372 hg58d07.x
11	83	76.9	510 10	AM0594372 hg58d07.x
12	82.4	76.3	520 10	AM0594372 hg58d07.x

Result No.	Score	Match length	DB ID	Description
13	82.2	76.1	529 10	AM0594372 hg58d07.x
14	82	75.9	529 10	AM0594372 hg58d07.x
15	82	75.9	547 10	AM0594372 hg58d07.x
16	81.8	75.7	291 10	AM0594372 hg58d07.x
17	81.4	75.4	408 10	AM0594372 hg58d07.x
18	81	75.0	176 10	AM0594372 hg58d07.x
19	80.4	74.4	366 10	AM0594372 hg58d07.x
20	79.8	73.9	347 10	AM0594372 hg58d07.x
21	79.4	73.5	474 10	AM0594372 hg58d07.x
22	79.2	73.3	588 10	AM0594372 hg58d07.x
23	78.8	73.0	448 10	AM0594372 hg58d07.x
24	78.8	73.0	521 10	AM0594372 hg58d07.x
25	78.4	72.6	593 10	AM0594372 hg58d07.x
26	78	72.2	440 10	AM0594372 hg58d07.x
27	78	72.2	562 10	AM0594372 hg58d07.x
28	77.4	71.7	478 10	AM0594372 hg58d07.x
29	77	71.3	356 10	AM0594372 hg58d07.x
30	77	71.3	397 10	AM0594372 hg58d07.x
31	77	71.3	407 10	AM0594372 hg58d07.x
32	77	71.3	507 10	AM0594372 hg58d07.x
33	77	71.3	507 10	AM0594372 hg58d07.x
34	76.8	71.1	239 11	AM0594372 hg58d07.x
35	76.4	70.7	973 11	AM0594372 hg58d07.x
36	76	70.4	143 10	AM0594372 hg58d07.x
37	76	70.4	388 10	AM0594372 hg58d07.x
38	76	70.4	670 11	AM0594372 hg58d07.x
39	75.8	70.2	518 10	AM0594372 hg58d07.x
40	75.4	69.8	519 10	AM0594372 hg58d07.x
41	75	69.4	379 10	AM0594372 hg58d07.x
42	75	69.4	466 10	AM0594372 hg58d07.x
43	75	69.4	467 10	AM0594372 hg58d07.x
44	75	69.4	495 10	AM0594372 hg58d07.x
45	75	69.4	509 10	AM0594372 hg58d07.x

ALIGNMENTS

RESULT 1
AM022375 219 bp mRNA EST 13-SEP-1999
LOCUS df37902.y1 Morton Fetal Cochlea Homo sapiens CDNA clone
DEFINITION IMAGE:2485802 5', mRNA sequence.
ACCESSION AM022375.1 GI:5675905
VERSION AM022375.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
Isolation of novel and known genes from a human fetal cochlear CDNA library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)
95130111

JOURNAL MEDLINE

CONTACT: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: ccmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC); see
http://www.nisc.nih.gov/.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Plate: LLM6185 row: M column: 3
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers

```

source
1. .219
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2485802"
/clone_lib="Morton Fetal Cochlea"
/issue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/organism="ear: Vector: pbluescript SK-; Site 1: EcorI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT      93 a      28 c      32 g      66 t
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 219;
Best Local Similarity 99.0%; Pred. No. 0.4;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggc 64
|||||
DB 113 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 172

QY 65 ttccaaaagaaaaaataaaaaaataaaaaa 103
|||||
DB 173 TTCCAAAAGAAAAAATAAAAAAATAAAAAA 211

RESULT 2
LOCUS      AM022874      392 bp      mRNA      EST      13-SEP-1999
DEFINITION  d145804.y1 Morton Fetal Cochlea Homo sapiens cDNA clone
IMAGE:2486430 5', mRNA sequence.
ACCESSION  AM022874
VERSION    AM022874.1 GI:5876404
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS    1 (bases 1 to 392)
Robertson, N.G., Khetarpal, U., Gutierrez-Espeleta, G.A., Bieber, F.R.
and Morton, C.C.
JOURNAL    Isolation of novel and known genes from a human fetal cochlear cDNA
MEDLINE    library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)
95130111
COMMENT     Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: comorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Plate: L1AM6187 row: G column: 7
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1. .392
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2486430"
/clone_lib="Morton Fetal Cochlea"
/issue_type="cochlea"

```

```

/issue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/organism="ear: Vector: pbluescript SK-; Site 1: EcorI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT      139 a      58 c      66 g      127 t      2 others
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 392;
Best Local Similarity 99.0%; Pred. No. 0.26;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggc 64
|||||
DB 261 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 320

QY 65 ttccaaaagaaaaaataaaaaaataaaaaa 103
|||||
DB 321 TTCCAAAAGAAAAAATAAAAAAATAAAAAA 359

RESULT 3
LOCUS      AM955633      447 bp      mRNA      EST      01-JUN-2000
DEFINITION  EST367703 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION  AM955633
VERSION    AM955633.1 GI:8145316
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS    1 (bases 1 to 447)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
JOURNAL    Assessment of gene expression patterns in a model of colon tumor
MEDLINE    metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
COMMENT     Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johng@tigr.org
Plate: 81
Seq primer: Reverse.
Location/Qualifiers
1. .447
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/issue_type="Vector: pbluescriptSKm"

BASE COUNT      184 a      69 c      64 g      130 t
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 447;
Best Local Similarity 99.0%; Pred. No. 0.24;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggc 64
|||||
DB 180 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 239

QY 65 ttccaaaagaaaaaataaaaaaataaaaaa 103
|||||
DB 240 TTCCAAAAGAAAAAATAAAAAAATAAAAAA 278

```


RESULT 4
LOCUS N68114/c
DEFINITION yz24a12.s1 Morton fetal Cochlea Homo sapiens CDNA clone
IMAGE:286846.3' similar to gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN
PRECUSOR (HUMAN);, mRNA sequence.
ACCESSION N68114
VERSION N68114
KEYWORDS GI:1224275
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
'B., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Mairra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
TITLE JOURNAL MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 291.
Location/Qualifiers
1. 409
/organism="Homo sapiens"
/db_xref="GDB:3893965"
/db_xref="taxon:9606"
/clone="IMAGE:286846"
/clone_lib="Morton Fetal Cochlea"
/tissue_type="Cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;
Site: 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGT 3'

BASE COUNT 143 a 61 c 49 g 155 t 1 others
ORIGIN

Query Match 79.1%, Score 85.4; DB 11; Length 409;
Best Local Similarity 98.0%; Pred. No. 0.38;
Matches 97; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgatattgagatgtttaataaa-tgtgaataaataaataaagcattgtgt 64
|||||
DB 107 TTTGATATGTGAGATGTTTAATTAATTGTAATAAATAAAGCATGTTGTT 48
|||||

OY 65 ttccaaagaagaaaaaataaataaataaataaataa 103
|||||
DB 47 TTCACAAAAAATAAATAAATAAATAAATAAATAA 9
|||||

RESULT 5
LOCUS AM069257/c
DEFINITION cr42c12.x1 Jia bone marrow stroma Homo sapiens CDNA clone

ACCESSION HBMS-Cr42c12.3', mRNA sequence.
VERSION AM069257
KEYWORDS AM069257.1 GI:6024178
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS Jia, L.B., Young, M.F., Touchman, J.W., Bouffard, G.G.,
Beckstrom-Sternberg, S.M., Green, E.D., Powell, J.I., Yang, L.M., Robey,
P.G., Hotchkiss, R.N. and Francomano, C.A.
SCAP: The Skeletal Genome Anatomy Project
Unpublished (1997)
Contact: libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@helix.nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).
Plate: 42 row: C column: 12
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
1. 591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HBMSC-Cr42c12"
/clone_lib="Jia bone marrow stroma"
/sex="mixed"
/tissue_type="bone marrow stroma"
/dev_stage="mixed"
/lab_host="XLI-Blue MRF/SOLR"
/note="Vector: pBluescript; Site: 1: EcoRI; Site: 2: XhoI;
mRNA made from human bone marrow stroma, CDNA made by
oligo-dt priming. Directionally cloned. Size-selected for
average insert size >0.5 kb. Library supplied by Dr. Libin
Jia (NHCRI)."

BASE COUNT 192 a 85 c 86 g 209 t 19 others
ORIGIN

Query Match 79.1%, Score 85.4; DB 10; Length 591;
Best Local Similarity 98.0%; Pred. No. 0.29;
Matches 97; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgatattgagatgtttaataaa-tgtgaataaataaataaagcattgtgt 64
|||||
DB 99 TTTGATATGTGAGATGTTTAATTAATTGTAATAAATAAAGCATGTTGTT 40
|||||

OY 65 ttccaaagaagaaaaaataaataaataaataaataa 103
|||||
DB 39 TTCACAAAAAATAAATAAATAAATAAATAAATAA 1
|||||

RESULT 6
LOCUS AM022830
DEFINITION df44f07.v1 Morton Fetal Cochlea Homo sapiens CDNA clone
IMAGE:2486557.5', mRNA sequence.
ACCESSION AM022830
VERSION AM022830
KEYWORDS AM022830.1 GI:5876360
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS Robertson, N.G., Khetarpal, U., Gutierrez-Espeleta, G.A., Bieder, F.R.
and Morton, C.C.
Isolation of novel and known genes from a human fetal cochlear CDNA

JOURNAL
MEDLINE
COMMENT

library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)

Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996

Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center. (NISC; see
http://www.nisc.nih.gov).

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Plate: LLAM6186 row: L column: 14
Seq primer: M13RPL reverse primer (ABI).

FEATURES
Source

Location/Qualifiers

1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2486557"
/clone_lib="Morton Fetal Cochlea"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ear; Vector: pBluescript SK-; Site: 1; EcorI:
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
378 of inserts <0.5 kb, 568 0.5-1.0 kb, 78 >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT 151 a 61 c 84 g 155 t 1 others
ORIGIN

Query Match 78.3%; Score 84.6; DB 10; Length 452;
Best Local Similarity 94.2%; Pred. No. 0.43;

Matches 98; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatagaataagcatgttgggt 64
|||||
Db 338 TTGTGATATGTGAGATGTTAAATTAATTTGAAAAAAGAAATGAACATGTTGGTT 397
QY 65 ttccaaaagaaaaaataaaaaaataaaaaaagcgcg 108
|||||
Db 398 TTCCAAAAGAAAAAATAAAAAAATAAAAAAAGCTCGGG 441

RESULT 7
LOCUS AI659779 184 bp mRNA EST 10-MAY-1999
DEFINITION t86c08.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2248430 3'
similar to gb:003464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
) ;, mRNA sequence.

ACCESSION AI659779
VERSION AI659779.1 GI:4763349
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 184)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

FEATURES
Source

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/db/ftp/image/image.html
Seq primer: -400P from Gibco.

Location/Qualifiers
1. 184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2248430"
/clone_lib="NCI-CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 57 a 27 c 20 g 80 t
ORIGIN

Query Match 78.1%; Score 84.4; DB 10; Length 184;
Best Local Similarity 98.0%; Pred. No. 0.85;

Matches 96; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatagaataagcatgttgggt 64
|||||
Db 98 TTGTGATATGTGAGATGTTAAATTAATTTGAAAAAAGAAATGAACATGTTGGTT 39
QY 65 ttccaaaagaaaaaataaaaaaataaaaaaagcgcg 102
|||||
Db 38 TTCCAAAAGAAAAAATAAAAAAATAAAAAAAGCTCGGG 1

RESULT 8
LOCUS AW594372 433 bp mRNA EST 22-MAR-2000
DEFINITION hg68d07.x1 NCI-CGAP_G6 Homo sapiens cDNA clone IMAGE:2950765 3'
similar to gb:003464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
) ;, mRNA sequence.

ACCESSION AW594372
VERSION AW594372.1 GI:7281630
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 433)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco
High quality sequence stop: 423.

FEATURES

Location/Qualifiers
1. 433

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2950765"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 159 a 71 c 51 g 152 t
ORIGIN

Query Match 77.6%; Score 83.8; DB 10; Length 433;
Best Local Similarity 97.0%; Pred. No. 0.54;
Matches 96; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 6 ttgtatattgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 64
|||||
Db 107 TTTGTATATGTCAGATGTTTAAATTAATGTCAGAAAAAATGAATGAACATGTTGGTT 48
|||||
OY 65 ttccaaagaagaaaaaataaaaaaataaaaaa 103
|||||
Db 47 TTCCAAAGAAAAAATAAATAAATAAAGAAAA 9
|||||

RESULT 9 A1445464 543 bp mRNA EST 13-APR-1999
LOCUS A1445464/c t344h05.x1 NCI_CGAP_Panl Homo sapiens cDNA clone IMAGE:2143449 3'
DEFINITION similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
); mRNA sequence.
ACCESSION A1445464
VERSION A1445464.1 GI:4288395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: c9apbs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbfp/image/image.html
Insert length: 1285 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 125.
Location/Qualifiers

FEATURES

source
1. 543

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2143449"
/clone_lib="NCI_CGAP_Panl"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 184 a 89 c 74 g 196 t
ORIGIN

Query Match 77.6%; Score 83.8; DB 10; Length 543;
Best Local Similarity 97.0%; Pred. No. 0.46;
Matches 96; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 6 ttgtatattgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 64
|||||
Db 103 TTTGTATATGTCAGATGTTTAAATTAATGTCAGAAAAAATGAATGAACATGTTGGTT 44
|||||
OY 65 ttccaaagaagaaaaaataaaaaaataaaaaa 103
|||||
Db 43 TTCCCAAGAAAAAATAAATAAATAAAGAAAA 5
|||||

RESULT 10 B1339163 275 bp mRNA EST 30-JUL-2001
LOCUS B1339163 364014 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION B1339163 B1339163.1 GI:15032446
ACCESSION B1339163.1 GI:15032446
VERSION EST.
KEYWORDS pig.
SOURCE Sus scrofa
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 275)
Fahrenkrug, S.C., Fieking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
JOURNAL Contact: Smith TPL
COMMENT USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGCTACGACGAC
Plate: 100 row: O column: 15
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

FEATURES

source
1. 275

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 104 a 47 c 43 g 80 t
ORIGIN

Query Match 77.4%; Score 83.6; DB 11; Length 275;
Best Local Similarity 93.3%; Pred. No. 0.78;
Matches 98; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

OY 6 ttgtatattgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 63
|||||
Db 104 TTTGTATATGTCAGATGTTTAAATTAATGTCAGAAAAAATGAATGAACATGTTGGTT 163
|||||

QY 64 ttcccaagaagaaaaaagcgcg 108
|||||
Db 164 TTTCAGAAAGAAAAAAGG 208

RESULT 11
AI620752 510 bp mRNA EST 15-DEC-1999
LOCUS AI620752/c
DEFINITION similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN)
ACCESSION AI620752
VERSION AI620752.1 GI:4629878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 510)
Natl-GAP http://www.ncbi.nlm.nih.gov/ncigap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 1768 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 303
POLYA-No.

FEATURES
source Location/Qualifiers
1..510
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="NCI-CGAP-2258927"
/clone_lib="NCI-CGAP-Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 170 a 90 c 73 g 177 t
ORIGIN

Query Match 76.3%; Score 83; DB 10; Length 510;
Best Local Similarity 94.2%; Pred. No. 0.58;
Matches 97; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 2 cccattgtatagtgttaataaa-tgtgaataatgaataagcatgtt 60
|||||
Db 111 CTTTGTGATAGTGAAGATTAAATCTGAAAAAATGAATTAACATGTTT 52

QY 61 ggtttccaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
Db 51 GGTITTCACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 9

RESULT 12
AI435232 520 bp mRNA EST 30-MAR-1999
LOCUS AI435232/c
DEFINITION similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN)

ACCESSION AI435232
VERSION AI435232.1 GI:4301740
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 520)
Natl-GAP http://www.ncbi.nlm.nih.gov/ncigap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 3135 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 280.

FEATURES
source Location/Qualifiers
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2130165"
/clone_lib="NCI-CGAP-Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 177 a 90 c 81 g 172 t
ORIGIN

Query Match 76.3%; Score 82.4; DB 10; Length 520;
Best Local Similarity 97.9%; Pred. No. 0.66;
Matches 94; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6 ttctatcgtgagatgttcaataaa-tgtgaataatgaataagcatgtt 64
|||||
Db 96 TTCTATATGTGAGATGTTTAAATTTGTGAAAAAATGAATTAACATGTTT 37

QY 65 ttccaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 100
|||||
Db 36 TTCCAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1

RESULT 13
AI697004 529 bp mRNA EST 17-DEC-1999
LOCUS AI697004/c
DEFINITION similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN)
ACCESSION AI697004
VERSION AI697004.1 GI:4984904
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 529)
Natl-GAP http://www.ncbi.nlm.nih.gov/ncigap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbip/image/image.html

Insert Length: 774 Std Error: 0.00
Seq primer: -40UP from Gidco
High quality sequence stop: 284.

FEATURES

source

Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2324602"
/clone_lib="NCI CGAP, Paul"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 171 a 93 c 78 g 187 t
ORIGIN

Query Match 76.1%; Score 82.2; DB 10; Length 529;
Best Local Similarity 96.0%; Pred. No. 0.69;
Matches 95; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttgggt 64
|||||
Db 102 TTTGCATATGTCAGATGTTTAAATGTCGAAATAATGAATGAATGATGTTGGTT 43
|||||
OY 65 ttccaaagaaaaaataaaaaataaaaaa 103
|||||
Db 42 TTCCCAAAAAAAAAAAAAAAAAAAAAA 4

RESULT 14

AM958385 529 bp mRNA EST 01-JUN-2000
LOCUS AM958385
DEFINITION EST370455 MAGE resequences, MAGE Homo sapiens CDNA, mRNA sequence.
ACCESSION AM958385
VERSION AM958385.1 GI:8148069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 529)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,
J. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Contact: John Quackenbush
COMMENT Unpublished (2000)
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 119
Seq primer: Reverse.

FEATURES

source

Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 178 a 79 c 90 g 173 t 9 others
ORIGIN

Query Match

75.9%; Score 82; DB 10; Length 529;

Best Local Similarity 95.9%; Pred. No. 0.72;
Matches 93; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttgggt 64
|||||
Db 317 TTTGCATATGTCAGATGTTTAAATGTCGAAATAATGAATGAATGATGTTGGTT 376
|||||
OY 65 ttccaaagaaaaaataaaaaataaaaaa 101
|||||
Db 377 TTCCAAANGAAAAAAAAAAAAAAAAAAAAA 413

RESULT 15

AU120039 547 bp mRNA EST 19-OCT-2000
LOCUS AU120039
DEFINITION HEMBA1 Homo sapiens CDNA clone HEMBA1007207 5', mRNA
sequence.
ACCESSION AU120039
VERSION AU120039.1 GI:10935274
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 547)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Salto, K., Kawai, Y.,
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.
TITLE HRI human CDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human CDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; CDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source

Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1007207"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pMD18SFL3"

BASE COUNT 180 a 84 c 93 g 185 t 5 others
ORIGIN

Query Match 75.9%; Score 82; DB 10; Length 547;
Best Local Similarity 97.9%; Pred. No. 0.7;
Matches 93; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttgggt 64
|||||
Db 287 TTTGCATATGTCAGATGTTTAAATGTCGAAATAATGAATGAATGATGTTGGTT 346
|||||
OY 65 ttccaaagaaaaaataaaaaataaaaaa 99
|||||
Db 347 TTCCAAANGAAAAAAAAAAAAAAAAAAAAA 381

Search completed: January 17, 2002, 06:42:50
Job time: 3077 sec

```

xx 18-APR-1996; 96US-0634546.
xx (TEXA ) UNIV TEXAS SYSTEM.
xx
xx PA Singh P, Wood TG;
xx PI
xx WPI; 2001-136591/14.
xx DR P-PSDB; AAV72384.
xx
xx PT Isolated antisense polynucleotide genetic construct for treatment of
xx colon cancer -
xx PS
xx Claim 1; Column 41-42; 35pp; English.
xx
xx CC The present invention relates to methods for the treatment of colon
xx cancer. The method involves inhibiting gastrin expression in
xx CC colon cancer cells using antisense construct.
xx CC The present sequence is a human preprogastrin cDNA used in the
xx CC construction of antisense construct. Gastrin is a peptide hormone
xx CC that plays a role in the initiation of colon tumours.
xx
xx SQ Sequence 383 BP, 89 A; 120 C; 111 G; 63 T; 0 other;
xx
xx
xx Query Match 50.1%; Score 52.6; DB 22; Length 383;
xx Best Local Similarity 82.8%; Pred No.1.2e-09;
xx Matches 72; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
xx
xx QY 19 cctgcagttctgtgtaacatgatgcagcagcatgtgtatgtctcttgcact 78
xx ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
xx Db 24 cctccacgctctgcagcagcagatgatgcagcagctatgtgtatgtcgtatcttgcact 83
xx ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
xx QY 79 aggatcaggcgccctctctgaagctt 105
xx || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
xx Db 84 -ggctctgycgcctctctgaagctt 109
xx
xx
xx RESULT 2
xx ID AAN90234 standard; DNA; 829 BP.
xx
xx AC AAN90234;
xx
xx DT 01-NOV-1989 (first entry)
xx
xx DE Gastrin recombinant DNA.
xx
xx KW Recombinant DNA; promoter; signal peptide; gastrin; shuttle
xx vector; yeast; Escherichia coli.
xx
xx OS Synthetic.
xx
xx FH Key Location/Qualifiers
xx FH exon 1..472
xx FT /tag= a
xx FT intron 473..602
xx FT exon /tag= b
xx FT /tag= c
xx FT misc_feature 78..81
xx FT /tag= d
xx FT misc_feature 288..291
xx FT /tag= e
xx FT misc_feature 320..325
xx FT /tag= f
xx FT misc_feature 436..441
xx FT /tag= g
xx FT misc_feature 711..716
xx FT /tag= h
xx FT misc_feature 768..771
xx FT /tag= i
xx FT misc_feature 794..797
xx

```

```

FT      /*tag= j
ET      262..694
ET      /*tag= k
XX
XX      JP64002579-A.
XX
XX      03-MAY-1989.
XX
XX      24-JUN-1987;    87JP-0156745.
XX
XX      24-JUN-1987;    87JP-0156745.
XX
XX      PA      (AJIN ) AJINOMOTO KK.
XX
XX      WP1: 1989-176418/24.
XX      P-PSDB; AAP90331.
XX
XX      Recombinant DNA encoding gastrin - also comprising promoter and signal
PT      peptide coding regions, inserted into shuttle vector to produce
PT      gastrin in yeast or E.coli.
XX
XX      Claim; page 524 and fig 2; 10pp; Japanese.
XX
XX      The recombinant DNA effectively produces gastrin in high quantities.
CC      It is made by integrating DNA fragment encoding gastrin into a shuttle
CC      vector which can grow in yeast or E.coli, eg pMB2, CY37, PAT77, YEP51,
CC      YE52 or pJDE218. The DNA encoding signal peptide is, eg from human
CC      gastrin, Mfalpha1 (yeast secretion hormone), yeast invertase, killer
CC      toxin, yeast alkaline or acid phosphatase, human lysozyme, human alpha-
CC      amylase or human trypsin inhibitor. DNA encoding gastrin is from, eg
CC      pH53. See also AAP0331. Misc. feature d is Sau3A site, e is Sau3A
CC      site, f is HindIII site, g is PvuII site, h is HindIII site, i is Sau3A
CC      site, and j is Sau3A site. The CDS is minus the Intron.
CC      Bases 262-318 encode signal peptide, and bases 319-429 encode gastrin.
XX
XX      Sequence 829 BP; 199 A; 251 C; 212 G; 167 T; 0 other;
SQ
Query Match          45.3%; Score 47.6; DB 10; Length 829;
Best Local Similarity 92.4%; Pred. No. 9.ge-08;
Matches   61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
Oy      40 gatgcagcagcatgtgtaatctgctcttcacttgacatcaggccgctctctcg 99
        |||||||
Db      261 gatgcagcagcatagtgtatctgctctgcatcttgcact-ggctctggcgcgctctctcg 319
Oy      100 aaagctt 105
        |||||
Db      320 aaagctt 325
RESULT      3
AC      AAT97434
XX      ID      AAT97434 standard; cDNA; 613 BP.
XX
XX      AAT97434;
XX
XX      05-MAY-1998 (first entry)
XX
XX      Human gastrin gene.
XX
XX      Human; gastrin; inhibition; colon cancer; colorectal cancer;
KW      treatment; prophylaxis; ss.
XX
XX      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..210
FT
FT      /*tag= a
FT      /product= "gastrin"
FT      /note= "no stop codon given: if the coding sequence is
TT      extended to position 225, it ends in a stop
TT      codon and Gly-Arg-Cys is added to the end
```


FT of the protein in AAW31634"
XX M09738584-A1.
XX 23-OCT-1997.
XX 16-APR-1997; 97WO-US06528.
XX 18-APR-1996; 96US-0634546.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Singh P, Wood TG;
XX WPI: 1997-526130/48.
XX P-PSDB: AAW31634.
XX Inhibition of growth of colon cancer cells - by transfection of
PT anti:sense gastrin gene sequences
XX
XX Claim 12; Page 48-49; 62pp: English.
XX
XX A method has been developed of inhibiting the growth of gastrin-
CC expressing colon cancer cells. The method comprises reducing gastrin
CC expression of the cells by transfecting the cells to induce antisense
CC polynucleotides inhibiting gastrin expression in the cells. The present
CC sequence encodes human gastrin. A composition has also been developed
CC for treating colon cancer, which comprises an expression construct
CC comprising a promoter functional in eukaryotic cells and a
CC polynucleotide encoding a gastrin gene positioned antisense to and
CC under the control of the promoter. The composition is used for treating
CC colorectal cancer, especially in humans. Antibodies against gastrin are
CC non-specific and cross-react with cholecystokinin. Also, antibodies are
CC not available against all precursor forms of gastrin, and may often be
CC antigenic themselves. The present method provides an alternative that
CC can be used for the treatment and prophylaxis of colon-cancer cell
CC growth.
XX
XX Sequence 613 BP; 135 A; 181 C; 173 G; 124 T; 0 other;
SQ

Query Match 44.4%; Score 46.6; DB 18; Length 613;
Best Local Similarity 92.3%; Pred. No. 2.1e-07;
Matches 60; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 41 atgcagcagcatgtgtatgtctgtcttgcactagatcagcgcccttctcga 100
Db 1 atgcagcagcatgtgtatgtctgtcttgcact-ggctctgcgcgccttctcga 59
QY 101 agctt 105
Db 60 agctt 64

RESULT 4
AAD02451
ID AAD02451 standard; DNA; 613 BP.
XX
XX AAD02451;
XX
XX 24-APR-2001 (first entry)
XX
XX Human partial preprogastrin encoding DNA.
XX
XX Human; cytosstatic; antisense construct; preprogastrin; therapy;
KW colon cancer; tumour; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 1..210
FT /*tag= a
FT /product= "Partial preprogastrin peptide"

FT /note= "Does not include stop codon"
FT /partial
XX US6165990-A.
XX 26-DEC-2000.
XX 15-MAY-1998; 98US-0079372.
XX 18-APR-1996; 96US-0634546.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Singh P, Wood TG;
XX WPI: 2001-136591/14.
XX P-PSDB: AAY72377.
XX Isolated antisense polynucleotide genetic construct for treatment of
PT colon cancer -
XX
XX Disclosure: Column 35-36; 35pp: English.
XX
XX The present invention relates to methods for the treatment of colon
CC cancer. The method involves inhibiting gastrin expression in
CC colon cancer cells using antisense construct.
CC The present sequence is a human partial preprogastrin DNA. Gastrin is
CC a peptide hormone that plays a role in the initiation of colon tumours.
XX
XX Sequence 613 BP; 135 A; 181 C; 173 G; 124 T; 0 other;
SQ

Query Match 44.4%; Score 46.6; DB 22; Length 613;
Best Local Similarity 92.3%; Pred. No. 2.1e-07;
Matches 60; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 41 atgcagcagcatgtgtatgtctgtcttgcactagatcagcgcccttctcga 100
Db 1 atgcagcagcatgtgtatgtctgtcttgcact-ggctctgcgcgccttctcga 59
QY 101 agctt 105
Db 60 agctt 64

RESULT 5
AAH18005/C
ID AAH18005 standard; CDNA; 2857 BP.
XX
XX AAH18005;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17809.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX

[illegible]

PR		25-APR-2000;	2000US-05592317.
PR		09-JUL-2000;	2000US-0598042.
PR		19-JUL-2000;	2000US-0620312.
PR		03-AUG-2000;	2000US-0653450.
PR		14-SEP-2000;	2000US-0662191.
PR		19-OCT-2000;	2000US-0693036.
PR		29-NOV-2000;	2000US-0727344.
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang Z, Wehnman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPJ; 2001-442253/47.		
DR	P-PsDB; AAM39311.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Claim 1; SEQ ID NO 670; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC	immunosuppressant and cytoskeletal activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombotic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, Leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
SQ	Sequence 4495 BP; 1401 A; 945 C; 1040 G; 1109 T; 0 other;		
	Query Match	28.8%; Score 30.2; DB 22; Length 4495;	
	Best Local Similarity .69.5%; Pred. No. 0.35;		
	Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0		
OY	15 acatctgcagcgttcgctgaagatgatcgagcagcatgctgctgctctt 73 II IIIII IIIII III III III IIII IIII I IIIII		
Db	3983 ACTTCGTGAGTTCCTCTCGACATCAAGACTGCCTGGTGTTGTTCTTT 3925		
RESULT 7			
AI60253/c			
ID	AA160253 standard; cDNA; 4775 BP.		
XX			
XX	AA160253:		
XX			
DT	22-OCT-2001 (first entry)		
DE	Human polynucleotide SEQ ID NO 4242.		
XX			
KM	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KM	peripheral nervous system; neurotrophic; central nervous system; CNS;		
KM	Alzheimer's, Parkinson's disease, Huntington's disease, haemostatic;		
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KM	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukemia; ss.		
XX			
OS	Homo sapiens.		
NN	WO200153312-A1.		
DD	26-JUL-2001.		

XX	26-DEC-2000; 2000WO-US34263.
XX	
XX	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0596042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0634550.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
XX	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Demanac RT;
XX	
XX	WPI: 2001-442253/47.
DR	P-PSDB; AAM41097.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Claim 1; SEQ ID NO 4242; 10076bp; English.
XX	
XX	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 4775 BP; 1494 A; 999 C; 1122 G; 1160 T; 0 other;
XX	
XX	
XX	Query Match 28.8%; Score 30.2; DB 22; Length 4775;
XX	Best Local Similarity 69.5%; Pred. No. 0.36;
XX	Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0
OY	15 acatccgcgaattctgtgtacagatgcacgacgatgtgtgtatgtgcctctt 73
DB	3944 ACTTCTGTGAGCTTCTCTCTCAGATACACTCAAGAAGCTGCGTGTGTTGCTTT 3886
XX	
XX	RESULT 8
XX	AAH41086/C
ID	AAH41086 standard; CDNA; 6006 BP.
XX	
XX	AAH41086;
XX	
DT	29-AUG-2001 (first entry)
XX	
DE	Human calcium binding/actin crosslinking protein (CBACP) cDNA.
KW	Human; calcium-binding/actin cross-linking protein; CBACP; infertility;
KW	immunological disease; hypertension; renal insufficiency; infection;
KW	muscular dystrophy; malnutrition; neurotransmission disorder; cardiac;
KW	kidney disease; anaemia; heart disease; cancer; immunomodulatory;
KW	hypertensive; nephritic; antiviral; antibacterial; antihaemic;
KW	cytostatic; ss.
XX	
XX	Homo sapiens
OS	

[illegible]

XX 25-FEB-2000; 2000EP-0301439.
PF
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139617.
PR 22-JUN-1999; 990S-0139699.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.

PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147034.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149929.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.

```
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157733.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 25.9%; Score 27.2; DB 21; Length 1797;
Best Local Similarity 67.9%; Pred. No. 3.3;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Oy 30 tctgtacagatgacgacgagtgtgtatgtctctcttccttcacagatca 85
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1632 TGTGTTGCTGTTGCAGACATGTGAGTCACTCTCCTCCATGAGCTCTATGA 1577
```

RESULT 10
AAT60450
ID AAT60450 standard; cDNA: 7918 BP.

XX AAT60450;

DT 09-JUL-1997 (first entry)

DE Melanoma-associated chondroitin sulphate proteoglycan (MCSP) cDNA.

KW MCSP; melanoma-associated chondroitin sulphate proteoglycan;
KW tumour; sarcoma; glioblastoma; diagnosis; therapy; vaccine; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..6969

FT sig_peptide 1..87

FT mat_peptide 88..6966

FT misc_difference 7434

FT /tag= d

FT /note= "base 7434 is given as n in the specification"

```
XX XX
PN W09713855-A1.
XX 17-APR-1997.
PD 17-APR-1997.
XX 10-OCT-1995; 95WO-EP03988.
XX 10-OCT-1995; 95WO-EP03988.
PR 10-OCT-1995; 95WO-EP03988.
XX (NOVS ) NOVARTIS AG.
XX Pluschke G, Schmid P;
PI WPI: 1997-235891/21.
DR P-PSDB; AAM15566.
XX Melanoma-associated chondroitin sulphate proteoglycan - useful for
PT prophylaxis or treatment of melanoma-associated tumours
XX Claim 6; Page 31-48; 88pp; English.
```

Query Match 25.9%; Score 27.2; DB 18; Length 7918;
Best Local Similarity 67.9%; Pred. No. 5.2;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Oy 44 cagcagcagatggtgtatgtctgtcttccttcacagatcagcgccttcctg 99
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 594 cagtgatgatgtgcccctggtctctctggtccacactctgtgtccttcctg 649
```

RESULT 11
AA157889
ID AA157889 standard; cDNA: 8570 BP.

XX AA157889;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 92.

KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

OS W0200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR	25-APR-2000;	20000US-0552317.
PR	19-JUL-2000;	20000US-0598042.
PR	19-JUL-2000;	20000US-0620312.
PR	03-AUG-2000;	20000US-0634560.
PR	14-SEP-2000;	20000US-0662191.
PR	19-OCT-2000;	20000US-0693036.
PR	29-NOV-2000;	20000US-0727344.
XX		
PA	(HYSE-) HYSEQ INC.	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX		
DR	WPI: 2001-442253/47.	
DR	P-PSDB; AAM38733.	
PT		
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
PS	Claim 1; SEQ ID NO 92; 10078bp; English.	
XX		
CC	The invention relates to human nucleic acids (AA15798-AA161369) and	
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SQ	Sequence 8570 BP; 1594 A; 2711 C; 2648 G; 1617 T; 0 other;	
	Query Match	25.9%; Score 27.2; DB 22; Length 8570;
	Best Local Similarity	67.9%; Pred. No. 5.3;
	Matches	38; Conservative 0; Mismatches 18; Indels 0; Gaps 0.
DY	44 cagcagcatgttgatgctgccttccttgactagatcagcgcccttcctg 99	
Db	957 cagtgatgatgtggccctgcctctctcgggccaccattctgctgcttcctg 1012	
	RESULT 12	
	AA159675/c	
ID	AA159675 standard; cDNA: 8689 BP.	
XX		
AC	AA159675:	
XX		
DT	22-OCT-2001 (first entry)	
DE	Human polynucleotide SEQ ID NO 3664.	
XX		
KM	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KM	peripheral nervous system; neuropathy; central nervous system; CNS;	
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;	
KM	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KM	leukemia; ss.	
XX		
OS	Homo sapiens.	
PN	WO200153312-A1.	
XX		
DD	26-JUL-2001.	

XX	26-DEC-2000;	2000MO-US34263.	
XX	21-JAN-2000;	2000US-0488725.	
XX	25-APR-2000;	2000US-0552317.	
PR	09-JUL-2000;	2000US-0598042.	
PR	19-JUL-2000;	2000US-0620312.	
PR	03-AUG-2000;	2000US-0653450.	
PR	14-SEP-2000;	2000US-0662191.	
PR	19-OCT-2000;	2000US-0693036.	
PR	29-NOV-2000;	2000US-0727344.	
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
XX	WPI: 2001-442253/47.		
DR	P-PSDB: AAM40519.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Claim 1; SEQ ID NO 3664; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
XX			
SO	Sequence 8689 BP; 1638 A; 2669 C; 2755 G; 1627 T; 0 other;		
	Query Match	25.9%; Score 27.2; DB 22; Length 8689;	
	Best Local Similarity	67.9%; Pred. No.5.3;	
	Matches	38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
QY	44 cagcagcagatgtgtatagtctgctcttgacatagatcagcgcctctctg 99		
DB	7614 CAGTGATGATGGGCCCTGGCTCTCTGGGCCCACACTCTGGCTGCCCTCCCTG 7559		
RESULT	13		
XX	AAF27661/C		
XX	AAF27661 standard; DNA: 2404 BP.		
AC			
AC	AAF27661;		
DT	30-MAR-2001 (first entry)		
XX			
DE	DNA encoding human transcriptional regulator protein #30.		
XX			
KW	Human; transcriptional regulator protein; TXREG; ds.		
XX			
OS	Homo sapiens.		
XX			
FN	WO200078954-A2.		
XX			
XD	28-DEC-2000.		
XX			


```

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3: SEQ ID 9245; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 501 BP; 145 A; 95 C; 86 G; 161 T; 14 other;

```

```

Query Match 25.3%; Score 26.6; DB 22; Length 501;
Best Local Similarity 64.4%; Pred. No. 3.7;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

```

OY 15 acatcctgcagttctgtgtaagatgagcagatgtgtgtatgtctgtctt 73
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 255 acttcctgtgagttctctgcacanaactgcaaaacccgcgtcgtggtgtcttctt 313

```

Search completed: January 17, 2002, 06:46:54
Job time: 3282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic --nucleic search, using sw model

Run on: January 17, 2002, 08:48:23 ; Search time 2206.96 Seconds
(without alignments)
807.308 Million cell updates/sec

Title: US-09-823-101-6
Perfect score: 108
Sequence: 1 acctattgtatattgagaga.....aaaaaaaaaaaaaagcgcg 108

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rtd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rtd:*
36: em_htg_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	69.4	343	11 G62185	G62185 EST544 Foet
2	61.8	57.2	1130	9 HUMCIA22	K01078 Human proco
3	61.8	57.2	38682	9 AF004877	AF004877 Homo sapi
4	61.8	57.2	141120	9 AC002528	AC002528 Human BAC
5	61.8	57.2	188781	2 AC022546	AC022546 Homo sapi
6	61.8	57.2	189401	2 AC092511	AC092511 Papio cyn
7	59.8	55.4	846	11 G15893	G15893 human STS C
8	59.8	55.4	5086	9 HSPPA2ICO	Z74616 H. sapiens m
9	59.8	55.4	5416	9 HUMCIA2	J03464 Human COL1A
10	56	51.9	236699	2 AC026891	AC026891 Mus muscu
11	54.2	50.2	1162	8 AF117707	AF117707 Lycopersi
12	54.2	50.2	3265	6 AX092338	AX092338 Sequence
13	53.6	49.6	4284	3 DDU20661	U20661 Dictyosteli
14	53.6	49.6	150844	33 AL357036	AL357036 Human DNA
15	53.6	49.6	169224	2 AC092814	AC092814 Homo sapi
16	53.6	49.6	193570	2 AL592297	AL592297 Homo sapi
17	53	49.1	542	3 PPNRSUB	X80129 P. falciparu
18	52.8	48.9	1052	3 AB070132	AB070132 Macaca fa
19	52.8	48.9	3499	9 HSU80919	U80919 Homo sapien
20	52.8	48.9	4678	9 HSU80917	U80917 Homo sapien
21	52.6	48.7	1804	5 TNU72071	U72071 Tilapia nil
22	52.6	48.7	2021	10 BC007159	BC007159 Mus muscu
23	52.6	48.7	2758	9 HSM802938	AL583915 Homo sapi
24	52	48.1	682	9 HST000239	AL360193 Homo sapi
25	51.6	47.8	5178	9 HSM802206	AL137716 Homo sapi
26	51.2	47.4	2802	9 AF084644	AF084644 Homo sapi
27	51.2	47.4	2905	9 AF084645	AF084645 Homo sapi
28	51	47.2	1064	5 AF081800	AF081800 Dictyoste
29	50.8	47.0	1064	5 AF067856	AF067856 Xenopus l
30	50.8	47.0	1658	10 AB030908	AB030908 Rattus no
31	50.8	47.0	2028	9 AB025432	AB025432 Homo sapi
32	50.6	46.9	2216	9 AK026788	AK026788 Homo sapi
33	50.4	46.7	2600	4 SSJ001202	AJ001202 Sus scrofa
34	50.2	46.5	3277	9 AF102546	AF102546 Homo sapi
35	50.2	46.5	3423	10 BC010717	BC010717 Mus muscu
36	50.2	46.5	234112	3 PPMALAP2	AL035675 Plasmodiu
37	50	46.3	934	3 DDU25660	U25660 Dictyosteli
38	50	46.3	1118	8 ATP22A	Y08781 A. thaliana
39	50	46.3	1310	3 AB013088	AB013088 Holotrich
40	50	46.3	2976	3 DDU09478	U09478 Dictyosteli
41	49.8	46.1	835	11 CNS06JGB	AL401553 T7 end of
42	49.8	46.1	1045	10 BC006754	BC006754 Mus muscu
43	49.8	46.1	3314	3 AF163834	AF163834 Dictyoste
44	49.6	45.9	1724	9 BC003101	BC003101 Homo sapi
45	49.6	45.9	2568	3 DDP8A7	X51947 Dictyosteli

ALIGNMENTS

RESULT 1
LOCUS G62185
DEFINITION EST544 Foetal cDNA library directionally cloned in Unizap XR vector
Egus caballus STS cDNA 3', sequence tagged site.
ACCESSION G62185
VERSION G62185
KEYWORDS G62185.1 GI:7677541
SOURCE STS.
ORGANISM horse.
Egus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE 1 (bases 1 to 343)
AUTHORS Brandon, R., Thomas, S. and Bell, K.
TITLE Development of equine expressed sequence tags for use in radiation
hybrid mapping
JOURNAL Unpublished (2000)
COMMENT

Contact: Richard Brandon
Australian Equine Blood Typing Research Laboratory

University of Queensland
St Lucia, Brisbane, Queensland, 4072, Australia
Tel: 61733658501
Fax: 61733654899
Email: brandone@plok.unq.edu.au
Primer A: GAGCATGTGTCGTGCTGACGAG
Primer B: ACCATGGGAAAGATGAAAC
STS size: 205
PCR Profile:

Cycle 1:

Cycles 2-31:

Cycle 32:

Thermal Cycler: MJ Research PTC-100
Protocol:

Template: 100ng
Primer: each 0.5 uM
dNTPs: each 200 uM
Taq Polymerase: 0.02 units/uL
Annealing Temp: 60
Total Vol: 20 uL

Buffer:
MgCl2: 1.5 mM
(NH4)2SO4: 16.6 mM
Tris-HCl: 67 mM
Gelatin: 200 ug/mL
Triton X-100: 0.45%

PCR works in the horse and not mouse for specific use in RH panel screening.

FEATURES
source Location/Qualifiers

1..343
/organism="Equus caballus"
/db_xref="taxon:9796"
/clone_lib="Foetal CDNA library directionally cloned in Unizap XR vector"
/tissue_type="60-day foetus"
/note="Vector: Bluescript; V-type: phagemid; A whole foetus was used as source of RNA in the construction of a stratagene Unizap XR directional library. The phagemid Bluescript was excised from the virus vector and used as template for sequencing reactions from the 3' end of the cDNA."

gene

1..343
/gene="COL1A2"
/note="collagen pro-alpha 2 (I)"
1..205
/gene="COL1A2"

STS

primer_bind

1..24
/gene="COL1A2"

primer_bind

complement(185..205)

BASE COUNT

118 a 48 c 57 g 119 t 1 others

ORIGIN

Query Match 69.4%; Score 75; DB 11; Length 343;

Best Local Similarity 98.9%; Pred. No. 0.0022; Mismatches 0; Indels 1; Gaps 1;

Matches 85; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 257 TTGTATATCTGAGATCTTTAAATTAATTTGAAAAAAGAAATTAAGCATGTTGGTT 316

QY 65 ttccaaagaaaaaataaaaaa 91

Db 317 TTCCAAAAGAAAAA 343

RESULT 2

HUMCIA22

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

CDS

Location/Qualifiers

1..1130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21.3-q22.1"
1..147
/gene="COL1A2"
/note="alpha-2 type I procollagen"
/codon_start=1
/db_xref="GDB:C00-119-062"
/protein_id="AA51887.1"
/db_xref="GI:179601"
/translation="KRTNEMGKTIIEFKTKNSALPFLDIAPLDIGADHFEFVDIG VCFK"

BASE COUNT 373 a 199 c 369 t
ORIGIN about 7.5 kb after <humcia21>.

Query Match 57.2%; Score 61.8; DB 9; Length 1130;

Best Local Similarity 90.6%; Pred. No. 0.23;

Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataa-tgtgaaaaaataagaaatgagatgttggtt 64

Db 918 TTGTATATCTGAGATCTTTAAATTAATTTGAAAAAAGAAATTAAGCATGTTGGTT 977

QY 65 ttccaaagaaaaaataaaaaa 89

Db 978 TTCCAAAAGAACATATTGACTTAAA 1002

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

2 (sites)

J. Biol. Chem. 258 (16), 10128-10135 (1983)

83290853

protein_bind 2487..2500
/gene="COL1A2"
/note="putative"
/citation=[7]
/bound_moiety="NF1"
2502..5143
/gene="COL1A2"
/citation=[11]
/citation=[7]
/number=1
3380..3386
/note="putative; bottom strand"
/bound_moiety="AP1"
3407..3413
/gene="COL1A2"
/note="putative"
/citation=[7]
/bound_moiety="AP1"
3716..3747
/citation=[7]
/rpl_type=tandem
/rpl_unit=gt
5144..5154
/gene="COL1A2"
/number=2
5155..5745
/gene="COL1A2"
/citation=[11]
/number=2
5746..5760
/gene="COL1A2"
/number=3
5761..6410
/gene="COL1A2"
/citation=[11]
/number=3

exon
intron

Query Match 57.2%; Score 61.8; DB 9; Length 38682;
Best Local Similarity 90.6%; Pred. No. 0.2;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtgtatctgagatgtttaataa-tgtgaaaaaatgaataagcatgttggct 64
|||||
Db 38471 TTGTATATGTGAGATGTAAATTTGTGAAAAAATGAATAAGCATGTGGT 38530
|||||

QY 65 ttccaaagaaaaaataaaaaa 89
|||||
Db 38531 TTCCAAAGAACATATGTGACTAAA 38555
|||||

RESULT 4
AC002528 141120 bp DNA PRI 04-FEB-2000
LOCUS Human BAC clone GSI-542D18 from 7q31-q32, complete sequence.
DEFINITION AC002528
AC002528
VERSION AC002528.1 GI:2388554
KEYWORDS HTG.
human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141120)
AUTHORS Kalkicki,J., Wamsley,P. and Gibson,A.
TITLE The sequence of H. sapiens BAC clone GSI-542D18
JOURNAL Unpublished
2 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission

JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H.GS542D18

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7 or send
mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc.
(http://www.genomesystems.com).
Cell line: Lymphoblastoid
Haplotypes: two
VECTOR: pBelOBAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is GSI-56H18, 200 bp overlap. The
actual start of this clone is at base position 1 of GSI-542D18;
actual end is at 141120 of GSI-542D18. This clone is part of an
unanchored island, orientation is unknown.

This clone contains STS's SMS53299 (NTD:91113722) and SMS5558
(NTD:9454686).

FEATURES
source
Location/Qualifiers
1..141120
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="GSI-542D18"
/clone_lib="GSBAC1"
/map="7q31-q32"
<680..24823
/gene="COL1A2"
join(<680..733,2277..2321,2613..2666,2766..2810,
3200..3253,3752..3850,3994..4038,4153..4251,4672..4725,
4850..4957,5319..5372,5486..5584,6498..6551,7014..7112,
7513..7566,8120..8173,8324..8377,8652..8705,9656..9700,
10835..10933,12154..12261,12929..12982,13928..13981,
14663..14716,14821..14874,14971..15078,15439..15492,
16329..16382,17387..17548,18766..18873,19547..19654,
20040..20093,20180..20287,20428..20481,20853..20960,
21438..21491,21618..21725,22057..22315,22723..22907,

```

23618..23860,24677..24823)
/gene="COLIA2"
/note="match to 274616 (NID:g1418929) (PID:g1418930): note
conflict at A.A. 1186 of (PID:1418930)."
/codon_start=1
/product="alpha2(i) collagen"
/protein_id="AAB69977.1"
/db_xref="GI:2386595"
/translation="GHNGLDGLKGGPGAGVKGEPGAGCENCTPGQTGARGLPGRRG
VGAPGAGARSDSGVGPVGPAGPTGSGAPGPPFPAPGKGTGAVGNAGPAPGPR
GEVGLPGISGPPGPPGPNAGNLTGAKGAGLPGVAGAPLPGPPGIPGPVAAATG
AKGLGEPPGAPSGKSGSKGSGASGAGGPGSGEGRKRNPNAGSAGPSPGL
KSGPSRGLPGADSGRAGVWGPGSGASGAPGAPRPNAGPAGPGLGSP
GNIGPAGKEGPGVGLPIDGRPPIGPAGARGPPGIGPPGPGKNDKDHAG
LAGAGGAPGDCNNGAOGPPGPGVGGKGGGPGGPGFGLPPSPAPGKGVGAV
RGLHEPGLPGPGRGRGPPGSGAAGPTPTISRPSGSPGPGDNGKGEVGVAV
GTAPSGSPGLPGEKAGITPGKKEKPEPLRGIIGMPRGAGAGAVAGAPAG
ATDGEAGAGAPGAPGAPRSGSPGREGVGPAGPAGPAGACQPGAKGGRGKGP
KGENGVGPTGPGVGAAGPAGPAGPAGSGPAGPAGPAGPAGPAGPAGPAGP
GPPGPPGAGKEGLRGRDGGVGTGTEVAGVPPGAGGEGSGEGAGTAGPPTPG
PGGLGAPGILGLPSRGERGLPGVAGVGEPPGLTAGPPGAPRPPGAVSGPNGA
PGEAGRDGNPNDRPPRDGPGHKGEGYPTONICPVAGAPGPHGVGPAGKHGR
GETPSGCVGAPGAVGPRPSGPGI RGDGKGPGRGLPLGLNGLGLPAG
HGDGAGGSGVGPAGPAGPAGPAGSGAGKDGRTGRTGVPAGIRGPGHGPAGP
PGPPGPGVSGGVDGFDGDFYRADQPRASPLRKYEDATLTKLNNOIETLTP
EGSRNPARTCRDLRLSHPEWSSGYWIDPNQGTMDAIFYCDPSTGETCIRAPDN
IPAKNWRSSKDKHVMLETINAGSOFEVNEGVTSKEMATOLAFMRLNAYSAQNT
TYHCKNSTAYMDEETGNLKAIVILGDSNDELVAAGNSRFYTYLVDCSKKTDMGK
TIEKTKNPSRLPFLDIAPLDIGADDEFVVDIGPVCFK"
1752..1793
/rpt_family="L1"
/note="match to EST 284030 (NID:g1806354)"
7513..7567
/gene="COLIA2"
/note="match to EST 284030 (NID:g1806354)"
8119..8174
/gene="COLIA2"
/note="match to EST 284030 (NID:g1806354)"
8321..8378
/gene="COLIA2"
/note="match to EST 284030 (NID:g1806354)"
repeat_region
/note="match to EST 284030 (NID:g1806354)"
complement(9477..9496)
/rpt_family="L1"
10173..10202
/rpt_family="L1"
complement(10280..10412)
/rpt_family="L1"
13919..13982
/gene="COLIA2"
/note="match to EST R46995 (NID:g807337)"
14662..14717
/gene="COLIA2"
/note="match to EST R46995 (NID:g807337)"
14673..14717
/gene="COLIA2"
/note="match to EST T11626 (NID:g596330)"
14821..14875
/gene="COLIA2"
/note="match to EST T11626 (NID:g596330)"
14821..14875
/gene="COLIA2"
/note="match to EST T11626 (NID:g596330)"
14971..15079
/gene="COLIA2"
/note="match to EST T11626 (NID:g596330)"
14971..15047
/note="match to EST T11626 (NID:g596330)"
15439..15493
/note="match to EST R46995 (NID:g807337)"
/gene="COLIA2"
/note="match to EST T11626 (NID:g596330)"

```

```

misc_feature
16241..16625
/gene="COLIA2"
/note="match to EST W93189 (NID:g1422342) zh45a06.r1"
misc_feature
complement(16379..16625)
/note="match to EST W93067 (NID:g1422239) zh45a06.s1"
misc_feature
complement(16570..16625)
/note="match to EST N69793 (NID:g1226373) yz60d05.s1"
misc_feature
complement(16646..17001)
/note="match to EST N69793 (NID:g1226373) yz60d05.s1"
16646..16959
/gene="COLIA2"
/note="match to EST H88548 (NID:g1070808) yw22a05.r1"
misc_feature
complement(16646..16796)
/note="match to EST W93067 (NID:g1422239) zh45a06.s1"
17445..17549
/gene="COLIA2"
/note="match to EST AA179691 (NID:g1761924) zp13g07.r1"
repeat_region
complement(18033..18324)
/rpt_family="AUU"
repeat_region
complement(18325..18416)
/rpt_family="L1"
misc_feature
18766..18877
/gene="COLIA2"
/note="match to EST AA179691 (NID:g1761924) zp13g07.r1"
19546..19655
/gene="COLIA2"
/note="match to EST AA179691 (NID:g1761924) zp13g07.r1"
20040..20094
/note="match to EST AA179691 (NID:g1761924) zp13g07.r1"
misc_feature
20178..20285
/gene="COLIA2"
/note="match to EST AA179691 (NID:g1761924) zp13g07.r1"

```

```

Query Match 57.2% Score 61.8; DB 9; Length 141120;
Best Local Similarity 90.6% Pred. No.0.19;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 6 ttgtatattgtagatgtttaataaa-tgtgaaaaaaatgaaataaagcatgtttg9t 64
Db 25587 TTTGTATATGTGAGATGTTTAAATAATGCAAAAAATGAAATTAAGCATGTTGCTT 25646
QY 65 ttccaaagaataaaaaa 89
Db 25647 TTCCAAAGAACAATATGAGTAA 25671

```

```

RESULT 5
AC022546 188781 bp DNA HTG 20-SEP-2000
LOCUS Homo sapiens chromosome 15 clone RP11-65B13 map 15, WORKING DRAFT
DEFINITION AC022546
ACCESSION AC022546.3 GI:10198375
VERSION AC022546.3
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188781)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Bocustavsky,L., Boukhalter,B., Brown,A., Bukett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehocck,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

```

TITLE
JOURNAL
COMMENT

Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, R., and Zody, M.

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced g1:9108379.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L1396
Center clone name: 65_B_13

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179493 bases at least Q40
Consensus quality: 183982 bases at least Q30
Consensus quality: 185689 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 186681; sum-of-contents
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contents

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

1      3528: contig of 3528 bp in length
*      3529 3628: gap of 100 bp
*      3629 5349: contig of 1721 bp in length
*      5350 5449: gap of 100 bp
*      5450 7196: contig of 1747 bp in length
*      7197 7296: gap of 100 bp
*      7297 9601: contig of 2305 bp in length
*      9602 9701: gap of 100 bp
*      9702 11931: contig of 2230 bp in length
*      11932 12031: gap of 100 bp
*      12032 14313: contig of 2282 bp in length
*      14314 14413: gap of 100 bp
*      14414 16778: contig of 2365 bp in length
*      16779 16878: gap of 100 bp
*      16879 19917: contig of 3039 bp in length
*      19918 20017: gap of 100 bp
*      20018 22727: contig of 2710 bp in length
*      22728 22827: gap of 100 bp
*      22828 28882: contig of 6055 bp in length
*      28883 28982: gap of 100 bp
*      28983 34336: contig of 5354 bp in length
*      34337 34436: gap of 100 bp
*      34437 40444: contig of 6008 bp in length
*      40445 40544: gap of 100 bp
*      40545 48669: contig of 8125 bp in length
*      48670 48769: gap of 100 bp
*      48770 57302: contig of 8533 bp in length
*      57303 57402: gap of 100 bp
*      57403 65250: contig of 7848 bp in length
*      65251 65350: gap of 100 bp

```

```

*      65351 89215: contig of 2365 bp in length
*      89216 89315: gap of 100 bp
*      89316 100430: contig of 11115 bp in length
*      100431 100530: gap of 100 bp
*      100531 115472: contig of 14942 bp in length
*      115473 115572: gap of 100 bp
*      115573 132563: contig of 16991 bp in length
*      132564 132663: gap of 100 bp
*      132664 152008: contig of 19345 bp in length
*      152009 152108: gap of 100 bp
*      152109 178451: contig of 26343 bp in length
*      178452 178551: gap of 100 bp
*      178552 188781: contig of 10230 bp in length.
*      Location/Qualifiers
1. 188781
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-65B13"
/clone_lib="RP11 Human Male BAC"
1. 3528
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
/note="assembly-fragment"
3629. 5349
/note="assembly-fragment"
5450. 7196
/note="assembly-fragment"
7297. 9601
/note="assembly-fragment"
9702. 11931
/note="assembly-fragment"
12032. 14313
/note="assembly-fragment"
14414. 16778
/note="assembly-fragment"
16879. 19917
/note="assembly-fragment"
20018. 22727
/note="assembly-fragment"
22828. 28882
/note="assembly-fragment"
28983. 34336
/note="assembly-fragment"
34437. 40444
/note="assembly-fragment"
40545. 48669
/note="assembly-fragment"
48770. 57302
/note="assembly-fragment"
57403. 65250
/note="assembly-fragment"
65351. 89215
/note="assembly-fragment"
89316. 100430
/note="assembly-fragment"
100531. 115472
/note="assembly-fragment"
115573. 132563
/note="assembly-fragment"
132664. 152008
/note="assembly-fragment"
152109. 178451
/note="assembly-fragment"
178552. 188781
/note="assembly-fragment"
vector_side:right"
clone_end:T7

```

```

BASE COUNT 56489 a 35268 c 32583 g 62340 t 2101 others
ORIGIN

```

Query Match 57.2% Score 61.8; DB 2; Length 188781;
 Best Local Similarity 90.6%; Pred. No. 0.19;
 Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgttgctt 64
 |||||||
 Db 82663 TTTGTATATGTGAGATGTTAAATAATTGTGAAAAAATAATGAATGAATGAATGTTGCTT 82722

OY 65 ttccaaagaaaaaataaaaaa 89
 |||||||
 Db 82723 TTCCAAAAGAACATATTGAGTAA 82747

RESULT 6
 AC092511 189401 bp DNA HNG 13-JUL-2001
 LOCUS Papio cynocephalus anubis clone RP41-444H21, WORKING DRAFT
 DEFINITION AC092511.1 GI:14717324
 VERSION AC092511.1
 KEYWORDS HNG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE olive baboon.
 ORGANISM Papio cynocephalus anubis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Papio.
 1 (bases 1 to 189401)
 Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
 Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L.,
 Granter, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,
 Lee-Jin, S.-O., Legaspi, R., Lim, M., Maduro, O.L., Maduro, V.B.,
 Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
 Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stancir, P., S.,
 Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L.,
 Walker, M.A., Wehrby, K.D., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 189401)
 Green, E.D.
 Direct Submission
 Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA

----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc_mouse@nhgri.nih.gov

----- Project Information
 Center project name: amn
 Center clone name: 444H21

----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; Version 0.990319
 Consensus quality: 189312 bases at least Q40
 Consensus quality: 189377 bases at least Q40
 Consensus quality: 189398 bases at least Q20
 Insert size: 154000; agarose-fp
 Insert size: 198000; pulse-field-gel
 Insert size: 189401; sum-of-contigs
 Quality coverage: 15.10x in Q20 bases; pulse-field-gel
 Quality coverage: 11.74x in Q20 bases; pulse-field-gel
 Quality coverage: 12.28x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1 189401: contig of 189401 bp in length.
 Location/Qualifiers
 1..189401
 /organism="Papio cynocephalus anubis"
 /db_xref="taxon:9535"
 /clone="RP41-444H21"
 /clone_1id="RP41"
 1..189401
 /note="assembly fragment"

BASE COUNT 60088 a 36311 c 35278 g 57724 t
 ORIGIN

Query Match 57.2% Score 61.8; DB 2; Length 189401;
 Best Local Similarity 90.6%; Pred. No. 0.19;
 Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgttgctt 64
 |||||||
 Db 96550 TTTGTATATGTGAGATGTTAAATAATTGTGAAAAAATAATGAATGAATGTTGCTT 96609

OY 65 ttccaaagaaaaaataaaaaa 89
 |||||||
 Db 96610 TTCCAAAAGAACATATTGAGTAA 96634

RESULT 7
 G15893
 LOCUS human STS CHLC_U01613_J03464.p56092 clone UTR_01613_J03464.
 DEFINITION G15893
 ACCESSION G15893
 VERSION G15893.1 GI:1161782
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human vector=pcDNA host=E.coli dut+ung+ (DH10B) Marker Selected
 genomic DNA prepared from XY individual of French nationality.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Metazoa; Chordata;
 Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
 Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
 Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 846)
 Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buettow, K.H.
 Cooperative Human Linkage Center
 Unpublished (1995)
 Synonyms: UTR_01613_J03464, CHLC_UTR_01613_J03464.T36142
 Contact: Dr. Jeffrey C. Murray
 UofI

The University of Iowa
 Department of Pediatrics, Iowa City, IA 52242, USA
 Tel: (319) 356-3508
 Fax: (319) 356-3347
 Email: jeff-murray@uiowa.edu

Primer A: TCACATGTTAGTGTGCGA
 Primer B: TTTGGCAATATATCTGCAT
 STS size: 247
 PCR Profile:
 denature: 30 seconds at 94 degrees C
 annealing: 75 seconds at 55 degrees C
 extension: 15 seconds at 72 degrees C
 PCR cycles: 27
 extension: 6 minutes at 72 degrees C

Protocol:
 Template: 30ng genomic DNA
 Primer: each 1.5 pmole
 dNTPs: each 200 uM
 Tag Polymerase: 0.3 units
 Total Vol: 10 uL

Buffer:
 MgCl2: 1.5mM
 KCl: 50mM
 Tris: 10mM

PH: 8.3
Prepared with primer pairs derived from J03464.
Location/Qualifiers
1. .846
/organism="Homo sapiens"
STS
Primer_bind 408. .654
Primer_bind complement(635. .654)
BASE COUNT 278 a 146 c 128 g 294 t
ORIGIN

Query Match 55.4%; Score 59.8; DB 11; Length 846;
Best Local Similarity 96.0%; Pred. No. 0.48;
Matches 72; Conservative 0; Mismatches 2; Indels 1; Caps 1;

Qy 6 ttctgatatgtgagatctttaaataaa-tgtgnaaaaaaagaataagcatgttgc 64
|||||
Db 771 TTTGTATATGAGATGTTTAAATTAATGTAAGAAAAAATGAATTAAGCATGTTGCTT 830
|||||

Qy 65 ttccaaaagaaaaa 79
|||||

Db 831 TTCCAAAGAACATA 845
|||||

RESULT 8
HSPPA2ICO 5086 bp mRNA PRI 07-MAR-1997
LOCUS H.sapiens mRNA for prepro-alpha2(I) collagen.
DEFINITION 274616
VERSION 274616.1 GI:1418929
KEYWORDS alpha2(I) collagen.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 54)
AUTHORS Dickson,L.A., de Wet,M., Di Liberto,M., Weil,D. and Ramirez,F.
TITLE Analysis of the promoter region and the N-propeptide domain of the human pro alpha 2(I) collagen gene
JOURNAL Nucleic Acids Res. 13 (10), 3427-3438 (1985)
MEDLINE 85242047
REFERENCE 2 (bases 55 to 64)
AUTHORS Sherwood,A.L., Bottenus,R.E., Martzen,M.R. and Bornstein,P.
TITLE Structural and functional analysis of the first intron of the human alpha 2(I) collagen-encoding gene
JOURNAL Gene 89 (2), 239-244 (1990)
MEDLINE 90323622
REFERENCE 3 (bases 65 to 2434)
AUTHORS Kuivaniemi,H., Tromp,G., Chu,M.L. and Prockop,D.J.
TITLE Structure of a full-length cDNA clone for the prepro alpha 2(I) chain of human type I procollagen. Comparison with the chicken gene confirms unusual patterns of gene conservation
JOURNAL Biochem. J. 252 (3), 633-640 (1988)
MEDLINE 88339824
REFERENCE 4 (bases 2435 to 3015; 4191 to 5086)
AUTHORS de Wet,M., Bernard,M., Benson-Chanda,V., Chu,M.L., Dickson,L., Weil,D. and Ramirez,F.
TITLE Organization of the human pro-alpha 2(I) collagen gene
JOURNAL J. Biol. Chem. 262 (33), 16032-16036 (1987)
MEDLINE 88058962
REFERENCE 5 (bases 3016 to 4190)
AUTHORS Makela,J.K., Vuorio,T. and Vuorio,E.
TITLE Growth-dependent modulation of type I collagen production and mRNA levels in cultured human skin fibroblasts
JOURNAL Biochim. Biophys. Acta 1049 (2), 171-176 (1990)
MEDLINE 90304220
REFERENCE 6 (bases 1 to 5086)
AUTHORS Dalglish,R.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) Raymond Dalglish, Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH,

United Kingdom
7 (bases 1 to 5086)
Dalglish,R.
TITLE The human type I collagen mutation database
JOURNAL Nucleic Acids Res. 25 (1), 181-187 (1997)
MEDLINE 97169389
FEATURES
source
1. .5086
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .139
1. .209
/number=1
140. .4240
/codon_start=1
/product="prepro-alpha2(I) collagen"
/protein_id="CAA9869.1"
/db_xref="GI:1418930"
/db_xref="SPTREMBL:Q15177"
/translation="MLSFVDTRTLILAVTLCLATCQSLQETVRKGPAGDGRNGRGER
GPPPPRGDGEDPTGPPPPGPPGLGNGNFAOYDGKGYLGPGMGLNGRGP
GAAGAPGPGFPGPAGPPEPGGTGPGAGRCGAPGPGKAGDEGHPKPPRPPRGVYG
PGCARGFPTGPIPGPKGTRHNGDLGKQGRAPVYKERRAPRNGNTPGOTGANGL
PGRGRVVGAPGAPGARSQSVGVPVGPAPIGSAGPPGPGAPRGELTGAAGNAGP
GAPRGVGEVGLPELISGVGPPEGNPANGLTGAKGAGLPBPVAGAPGLPPGRIPGVP
AAGATGARGLVGPVGPAGSKSGESGNGEBSAGPQGPSPGSEGRKPNGBAGSAGP
PGPPGLRSGSPGSKGLPGADGRAGVMPGSPGSRGAGVAGPAGPAGDAGREGEGLMPR
GLPSPGNIGPAGKEGVPGLPGIDGRPPITGPAGARFEPNGITGPPKPTDPPKNG
DKGHAGLAKARKAPRDGNNGAGPPGPGVGGKGEQGPAPPPGQGLPGSGAPGE
VGPGRGLHGERLPGPAPRPERKPPPESSAAGTGTGTGSRGSPGPPGDPNGNGE
GVVGAAGTAGPSCPSGLPGERGAGIIPGKGKGEGLGELIGNPDRDARAHKAVG
APGPAGATDGRGAGAACGAPGAPGSPGSEGEVGPAPNGFPAGPAGAAQPGAKGE
RAKGGKENGNGVPTGPGVGAAGPAGPNNPPGPGSRGSGGPGMTGPPGAGARTGP
GPGSGSPGPPGPPGKEGLRGRDGPQVGTGVEGAVGPPGPFAGEKPPSGEATGAG
PPTETPPQGLILAPGLIGLPGSKGEGGLPVAGAVGERPPLIAGPPRARGPGAVGS
PGVNGAPGAGARDGNDGPPRDGQPEHKKERKTPGNTIGVGAAGAPGPPGVPVGA
GKHNGNETGSPGVPVAGVPPRPSGPGIIRGDKGERGPRGLPGLKQHNGLG
LPGLAGHGDGAGSVGPAGPAGPAGPAGPAGRTGTPGTGVPAGTRGQKHGPP
AGPAGGPPGPPGPGVSGGYDFGYDDDFYRADOPRSPGLRKDYEDVTLISLNNQI
ETLTPPEGSRKNYARTCRDLRLSHPEWSSGYWIDPNOGCTMDAIIKYVDFTSGETCI
RQOPENIPAKNNYRSKDKKHVLTGTTINAGSQFENVNGVYSKEATDILAMRLLAN
VASONTTHCKNSIAIWDDETGKLAVALTQGSNDVELVAEGNSRFTYTVLVLDGCSKK
TNEWCKTILEYTNKPSRLPFLDIAPLDIGADHEFVDIGVCFK"
140. .205
206. .376
/note="N.propeptide"
210. .220
/number=2
221. .235
/number=3
236. .271
/number=4
272. .364
/number=5
365. .418
/number=6
377. .409
/note="N.telopeptide"
410. .3451
/note="triple_helix"
419. .463
/number=7
464. .517
/number=8
518. .571
/number=9
572. .625
/number=10
626. .679
/number=11
680. .733
/number=12
734. .778

exon /number-13
779..832
/number-14
833..877
exon /number-15
878..931
/number-16
932..1030
exon /number-17
1031..1075
/number-18
1076..1174
exon /number-19
1175..1228
/number-20
1229..1336
exon /number-21
1337..1390
/number-22
1391..1489
exon /number-23
1490..1543
/number-24
1544..1642
exon /number-25
1643..1696
/number-26
1697..1750
exon /number-27
1751..1804
/number-28
1805..1858
exon /number-29
1859..1903
/number-30
1904..2002
exon /number-31
2003..2110
/number-32
2111..2164
exon /number-33
2165..2218
/number-34
2219..2272
exon /number-35
2273..2326
/number-36
2327..2434
exon /number-37
2435..2488
/number-38
2489..2542
exon /number-39
2543..2704
/number-40
2705..2812
exon /number-41
2813..2920
/number-42
2921..2974
exon /number-43
2975..3082
/number-44
3083..3136
exon /number-45
3137..3244
/number-46
3245..3298
exon /number-47
3299..3406
/number-48
3407..3665
exon /number-49

misc_feature 3452..3496
/note="C_telopeptide"
misc_feature 3497..4237
/note="C_propeptide"
exon 3666..3850
/number=50
exon 3851..4093
/number=51
misc_feature 3938..3946

Query Match 55.4%; Score 59.8; DB 9; Length 5086;
Best Local Similarity 96.0%; Pred. No. 0.44;
Matches 72; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 6 ttgtatattgtgagatgtttaataaa-tgtgaataaataaagcatgtttggtt 64
|||||
Db 5011 ttgtatattgtgagatgtttaataaattgtgaaaaaaatgaataacatgtttggtt 5070
|||||

Oy 65 ttccaaagaataaa 79
|||||

Db 5071 ttccaaagaacacta 5085

RESULT 9

HUMCIA2 5416 bp mRNA PRI 31-OCT-1994
LOCUS Human collagen alpha 2 type I mRNA, complete cds, clone pHCOL2A1.
DEFINITION J03464 M18057 X02488
ACCESSION J03464.1 GI:179595
VERSION alpha-2 type I collagen; collagen.
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 751)
Dickson,L.A., de Wet,W., Di Liberto,M., Well,D. and Ramirez,F.
Analysis of the promoter region and the N-propeptide domain of the
human pro alpha 2(I) collagen gene
Nucleic Acids Res. 13 (10), 3427-3438 (1985)
85242047
2 (bases 335 to 5416)
de Wet,W., Bernard,M., Benson-Chanda,V., Chu,M.L., Dickson,L.,
Well,D. and Ramirez,F.
Organization of the human pro-alpha 2(I) collagen gene
J. Biol. Chem. 262 (33), 16032-16036 (1987)
88058962
3 (sites)
Wirtz,M.K., Gianville,R.W., Steinmann,B., Rao,V.H. and
Hollister,D.W.
Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids
comprising the N-telopeptide region of a pro-alpha 2(I) chain
J. Biol. Chem. 262 (34), 16376-16385 (1987)
88059013
4 (sites)
Well,D., Bernard,M., Combates,N., Wirtz,M.K., Hollister,D.W.,
Steinmann,B. and Ramirez,F.
Identification of a mutation that causes exon skipping during
collagen pre-mRNA splicing in an Ehlers-Danlos syndrome variant
J. Biol. Chem. 263 (18), 8561-8564 (1988)
88243702
5 (sites)
Marini,J.C., Lewis,M.B., Wang,Q., Chen,K.C. and Orrison,B.M.
Serine for glycine substitutions in type I collagen in two cases of
type IV osteogenesis imperfecta
J. Biol. Chem. (1992) in press
[2] revises [1].
[3] sites: Ehlers-Danlos syndrome type VIIB deletion. [4] sites:
polymorphism site.
Draft entry and computer-readable sequence [2] kindly provided by
F.Ramirez 18-OCT-1987.
EMBL features not translated to GenBank features:
key from 249 to 253
PRM pot. CANT-box


```

REFERENCE      3 (bases 1 to 1162)
AUTHORS       Nersisyan,A.M. and Valentine,J.S.
TITLE         Direct Submission
JOURNAL       Submitted (23-AUG-1999) Chemistry & Biochemistry, University of
                                           California Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095,
                                           USA
REMARK        Sequence update by submitter
COMMENT       On Aug 23, 1999 this sequence version replaced gi:4219091.
FEATURES
Source
transit_peptide <1..260
CDS
<1..935
/gene="CCS"
/note="similar to Saccharomyces cerevisiae Lys7 and Homo
sapiens CCS; complements aerobic lysine auxotrophy of
Saccharomyces cerevisiae Lys7 null strain"
/codon_start=3
/product="putative copper/zinc superoxide dismutase copper
chaperone precursor"
/protein_id="AADI2307.2"
/db_xref="GI:5759320"
/translation="AFIRSTVTAKTTAIAAIPAFAAVSSISSSOFERPIKLKRG
SISSNSIIISLFAKNIOKKSPSALIMETHSSNHQSSDNGVYLPPELTFFNVDKQ
OGCVSAKSKLQTYEGYKNDVDDIDNOVAFITLGSPKTMTEALEQGRKARLIGDCV
PDDPLISAAYAEFGPDIFGVRLAQNMELTRLEAFNSGSPGKHAMVSINEFDLIRF
GAASVTGLSYSLPLGDICTLDVDEGEFVGGPKELRVADLIGRAIVAVYTEDKSDPG
LTAIVAVIRSAVGAVENYKKICTCDGTIWEATSKI"
<1..1162
/gene="CCS"

```

		/product="putative copper/zinc superoxide dismutase copper	
BASE COUNT	393 a 200 c 248 g 321 t		
ORIGIN			
Query Match	50.2%;	Score 54.2;	DB 8; Length 1162;
Best Local Similarity	71.7%;	Pred. No. 3.4;	
Matches	71; Conservative	0; Mismatches	28; Indels 0; Gaps 0;
QY	5 attgtatattgtagtgctttaaatgtgaaataaataagcatgtttgctt 64		
Db	1003 ATGCTTTTAAGTTTGATGTCAGACGTAATGATCATCAAAATTAATGTAATCATTTGTTGCT 1062		
QY	65 ttccaaaagaaaaaataaaaaataaaaaaataaaaaa 103		
Db	1063 CTCAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1101		
RESULT 12			
AX092338			
LOCUS	AX092338	3265 bp	DNA
DEFINITION	Sequence 69 from Patent WO0116318.		
ACCESSION	AX092338		PAT
VERSION	AX092338.1	GI:13444483	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 3265)		
AUTHORS	Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A., Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and Wood,W.I.		
TITLE	Secreted and transmembrane polypeptides and nucleic acids encoding the same		
JOURNAL	Patent: WO 0116318-A 69 08-MAR-2001;		
FEATURES	location/Qualifiers		

XX AL357036.8
SV 27-MAY-2000 (Rel. 63, Created)
XX 23-JAN-2001 (Rel. 66, Last updated, Version 8)
DT
XX Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-565J7
XX
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
XX [1]
XX Plumb B.;
RL Submitted (21-JAN-2001) to the EMBL/GenBank/DBJ databases.
RL Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
RL humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
XX
XX ----- Genome Center
CC Center: Sanger Centre
CC Center code: SC
CC Web site: http://www.sanger.ac.uk
CC Contact: humquery@sanger.ac.uk
CC ----- Project Information
CC Center project name: ba56537
CC ----- Summary Statistics
CC Assembly program: XGAP4; version 4.5
CC Sequencing vector: plasmid; L08752; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Consensus quality: 14187 bases at least Q40
CC Consensus quality: 144912 bases at least Q30
CC Consensus quality: 146636 bases at least Q20
CC Insert size: 168730; 3.3% error; agarose-fp
CC Quality coverage: 4.18x in Q20 bases; sum-of-contigs
CC Quality coverage: 4.07x in Q20 bases; agarose-fp
CC
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 18 contigs. The true order of the pieces is
CC * not known and their order in this sequence record is
CC * arbitrary. Where the contigs adjacent to the vector can
CC * be identified, they are labelled with 'clone_end' in the
CC * feature table. Some order and orientation information
CC * can tentatively be deduced from paired sequencing reads
CC * which have been identified to span the gap between two
CC * contigs. These are labelled as part of the same
CC * 'fragment_chain', and the order and relative orientation
CC * of the pieces within a fragment_chain is reflected in
CC * this file. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence as
CC * soon as it is available and the accession number will be
CC * preserved.
CC
CC 1 13329 contig of 13329 bp in length
CC 1 13330 24463 contig of 11134 bp in length; fragment_chain 1
CC 24564 35124 contig of 10561 bp in length; fragment_chain 1
CC 35225 43383 contig of 8159 bp in length; fragment_chain 1
CC 43484 50759 contig of 7276 bp in length; fragment_chain 1
CC 50860 60970 contig of 10111 bp in length; fragment_chain 1
CC 61071 66063 contig of 4993 bp in length; fragment_chain 2
CC 66164 68816 contig of 3653 bp in length; fragment_chain 2
CC 69117 74097 contig of 4181 bp in length; fragment_chain 2
CC 74198 77759 contig of 3562 bp in length; fragment_chain 3
CC 77860 86114 contig of 8255 bp in length; fragment_chain 3
CC 86215 88174 contig of 11960 bp in length; fragment_chain 3
CC 88275 105363 contig of 7089 bp in length; fragment_chain 3
CC 105464 113505 contig of 9842 bp in length; fragment_chain 4
CC 113406 120735 contig of 5330 bp in length; fragment_chain 4
CC 120836 133308 contig of 12373 bp in length; fragment_chain 4
CC 133309 148679 contig of 15371 bp in length
CC

CC * 148780 150844 contig of 2065 bp in length
XX
XX Key Location/Qualifiers
FH source 1..150844
FT /chromosome="1"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="RP11-565J7"
FT /clone_11b="RPC1-11.2"
FT 1..13329
FT /note="assembly_fragment:01838"
FT /note="clone_end:SP6"
FT /note="vector_side:left"
FT 13330..24463
FT /note="assembly_fragment:01484"
FT /note="fragment_chain:1"
FT 24564..35124
FT /note="assembly_fragment:01776"
FT /note="fragment_chain:1"
FT 35225..43383
FT /note="assembly_fragment:01297"
FT /note="fragment_chain:1"
FT 43484..50759
FT /note="assembly_fragment:01491"
FT /note="fragment_chain:1"
FT 50860..60970
FT /note="assembly_fragment:01931"
FT /note="fragment_chain:1"
FT 61071..66063
FT /note="assembly_fragment:00043"
FT /note="fragment_chain:2"
FT 66164..69117
FT /note="assembly_fragment:00370"
FT /note="fragment_chain:2"
FT 69117..74097
FT /note="assembly_fragment:01845"
FT /note="fragment_chain:2"
FT 74198..77759
FT /note="assembly_fragment:01000"
FT /note="fragment_chain:3"
FT 77860..86114
FT /note="assembly_fragment:01842"
FT /note="fragment_chain:3"
FT 86215..98174
FT /note="assembly_fragment:01350"
FT /note="fragment_chain:3"
FT 98275..105363
FT /note="assembly_fragment:01525"
FT /note="fragment_chain:4"
FT 105464..115305
FT /note="assembly_fragment:00635"
FT /note="fragment_chain:4"
FT 115406..120735
FT /note="assembly_fragment:01073"
FT /note="fragment_chain:4"
FT 120836..133208
FT /note="assembly_fragment:00972"
FT 133309..148679
FT /note="assembly_fragment:01642"
FT 148780..150844
FT /note="assembly_fragment:00704"
FT /note="clone_end:T7"
FT /note="vector_side:right"
XX
SO Sequence 150844 BP; 44269 A; 29934 C; 30171 G; 44748 T; 1722 other;

Query Match 49.6%; Score 53.6; DB 33; Length 150844;
Best Local Similarity 77.4%; Pred. No. 3.5;
Matches 65; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Oy 20 atgtttaataatgtgaaataatgaataaagcagttgtgtttcccaaaagaaaaa 79

```

Db 79337 ATGCATTAATAGTGTATAAAAAAAAAAATAAGACCTCTGGCTTAATAAAAAAAAAA 79396
QY 80 aaaaaaaaaaaaaaaaaaaaaa 103
Db 79397 AAAAAAAAAAAAAAAAAAAAAA 79420

```

```

RESULT 15
AC092814/c DNA HTG 28-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-565J7, WORKING DRAFT SEQUENCE,
DEFINITION 3 unordered pieces.
ACCESSION AC092814.1 GI:15027767
VERSION AC092814 AL357036
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169224)
Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 169224)
Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
Haugen,E.D.
Direct Submission
Submitted (28-JUL-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jul 28, 2001 this sequence version replaced gi:12539675.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UMG
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-565J7 (sc0180)
----- Summary Statistics
Sequencing vector: plasmid; 46% of reads
Sequencing vector: plasmid; 108752; 54% of reads
Chemistry: Dye-terminator Big Dye; 89% of reads
Chemistry: Dye-terminator Big Dye; 11% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167317 bases at least Q40
Consensus quality: 168242 bases at least Q30
Consensus quality: 168768 bases at least Q20
Insert size: 169024; sum-of-ctrls
Quality coverage: 8.3x in Q20 bases; sum-of-ctrls

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 16209: contig of 16209 bp in length
* 16210 16309: gap of unknown length
* 16310 59803: contig of 43494 bp in length
* 59804 59903: gap of unknown length
* 59904 169224: contig of 109321 bp in length.
Location/Qualifiers
1. 169224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-565J7"
/clone_1db="RPC1 human BAC library 11"
FEATURES
source

```

```

misc_feature 1..16209
/note="assembly_name:Contig15"
misc_feature 16310..59803
/note="assembly_name:Contig16"
misc_feature 59904..169224
/note="assembly_name:Contig17"
BASE COUNT 48926 a 33490 c 34553 g 51952 t 303 others
ORIGIN

```

```

Query Match 49.6%; Score 53.6; DB 2; Length 169224;
Best Local Similarity 77.4%; Pred. No. 3.5;
Matches 65; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 20 atgtttaataatgtgaataaataaataaagcagttgtttccaaagaaataa 79
Db 53306 ATGCATTAATAGTGTATAAAAAAAAAAATAAGACCTCTGGCTTAATAAAAAAAAAA 53247
QY 80 aaaaaaaaaaaaaaaaaaaaaa 103
Db 53246 AAAAAAAAAAAAAAAAAAAAAA 53223

```

Search completed: January 17, 2002, 08:48:53
Job time: 10221 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:42:50 ; Search time 2644.98 Seconds
(without alignments)
426.564 Million cell updates/sec

Title: US-09-823-101-7

Perfect score: 105
Sequence: 1 tgcagtggtgcagacatcc.....ggccgcctctctgaagctt 105

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	31	29.5	545	13	AZ446318 1M0242M08
C 2	30.6	29.1	551	13	AQ499366 HS_5199_B
C 3	30.6	29.1	601	13	AZ415460 1M0190E03
C 4	30.2	28.8	260	11	D81803 HUM409B05B
C 5	30.2	28.8	331	11	H91634 yv03f12.s1
C 6	30.2	28.8	359	10	AM006058 w281d05.x
C 7	30.2	28.8	361	10	AA252861 zt76c01.s
C 8	30.2	28.8	386	10	BE223043 hu46d12.x
C 9	30.2	28.8	394	10	AA460473 zx61e07.r
C 10	30.2	28.8	402	10	AA130279 z129c07.r
C 11	30.2	28.8	404	11	R76040 y160c01.r1
C 12	30.2	28.8	416	10	AV725870 AV725870

Result No.	Score	Query Match	Length	DB ID	Description
C 13	30.2	28.8	418	10	A1076519 oz09f04.x
C 14	30.2	28.8	422	11	N90308 zb21d12.s1
C 15	30.2	28.8	422	11	N94771 yv61f02.r1
C 16	30.2	28.8	424	10	A1033293 ow74d10.s
C 17	30.2	28.8	429	10	A1200105 q190e09.x
C 18	30.2	28.8	432	11	w73949 zd71f109.s1
C 19	30.2	28.8	443	10	AA017343 ze52e08.r
C 20	30.2	28.8	447	10	A1275117 q170f02.x
C 21	30.2	28.8	448	10	A1131149 qc15b01.x
C 22	30.2	28.8	480	10	AA461041 zx61e07.s
C 23	30.2	28.8	486	10	A1955225 wx31e04.x
C 24	30.2	28.8	498	10	AA782083 a147d09.s
C 25	30.2	28.8	514	10	A1382006 te33d12.x
C 26	30.2	28.8	537	10	A1752305 c014h04.x
C 27	30.2	28.8	546	10	AM007337 ws51c11.x
C 28	30.2	28.8	561	10	BE002118 pml-BN008
C 29	30.2	28.8	582	10	A1798790 we92f11.x
C 30	30.2	28.8	625	10	AV658702 AV658702
C 31	30.2	28.8	679	10	AV731307 AV731307
C 32	30.2	28.8	699	10	AV731039 AV731039
C 33	30.2	28.8	721	10	AV731332 AV731332
C 34	30.2	28.8	938	10	AL573092 AL573092
C 35	30	28.6	755	13	CNS01P9L AL154537 Anopheles
C 36	29.6	28.2	933	13	CNS03WSJ AL264124 Tetraodon
C 37	29.4	28.0	508	10	AM634290 AM634290
C 38	29.4	28.0	534	10	AM460975 AM460975
C 39	29.2	27.8	559	10	AA019909 AA019909
C 40	29	27.6	471	13	AQ735408 HS_3024_B
C 41	28.8	27.4	456	12	AK019058 Mus muscu
C 42	28.8	27.4	456	12	AK008062 Mus muscu
C 43	28.8	27.4	456	12	AK008159 Mus muscu
C 44	28.8	27.4	456	12	AK008313 Mus muscu
C 45	28.8	27.4	456	12	AK008420 Mus muscu

ALIGNMENTS

RESULT 1
A2446318/c
LOCUS A2446318 545 bp DNA
DEFINITION 1M0242M08R Mouse 10kb plasmid U0GC1M library Mus musculus genomic
clone U0GC1M0242M08 R, DNA sequence.

ACCESSION A2446318
VERSION A2446318.1 GI:10597007
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00
Plate: 0242 row: M column: 08

Seq primer: CACACGAGAAACACCTATGACC
Class: plasmid ends
High quality sequence stop: 545.

FEATURES
SOURCE Location/Qualifiers
1..545

BASE COUNT
ORIGIN

163 a 150 c 120 g 112 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U08C1M0242M08"
/clone_1ib="Mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g147321141gb1Af129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

Query Match	Best Local Similarity	29.5%	Score 31	DB 13	Length 545
Matches 49	Conservative	0	Mismatches 30	Indels 0	Gaps
OY	5	ggtgtgcaggaatcctctccagttgtgtgaatagatgcgcgcagcatgtgtatgtg	64		
Db	480	gntttgcacgaacaccttgcagactttctgttgaatgctctgttgaagacagcttggttttg	422		
OY	65	ctgcctcttgcactagat	83		
Db	420	ctgctgtttcaccattgggt	402		

RESULT	2	LOCUS	551 bp	DNA	GSS	28-Apr-1999
		ACCESSION	AF0499366			
		VERSION	AF0499366.1	GI:4699478		
		KEYWORDS	GSS.			
		SOURCE	human.			
		ORGANISM	homo sapiens			
		REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. I (bases 1 to 551)			
		AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.			
		TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome			
		JOURNAL	Proc. Natl. Acad. Sci. U. S. A.	96 (17),	9739-9744	(1999)
		MEDLINE	99380589			
		COMMENT	Contact: Mahairas GG, Wallace JC, Hood L			

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPEC-11. For BAC library availability, please contact Plietser de Jong (plieterdejong.med.buffalo.edu). Clones may be purchased from

BACpac Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Researc h Genetics (<http://www.rhgsc.washington.edu>)
BAC end Web Server:
<http://www.rhgsc.washington.edu>
Plate: 775 .row: H column: 4
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 551.
Location/Qualifiers
1. .551

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate:775 Col=4 Row=H"
/clone_11b="RPCT-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen d and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT      143 a      167 c      115 g      124 t      2 others
ORIGIN

```

	Query Match	29.1%	Score 30.6	DB 13	Length 551
	Best Local Similarity	58.1%	Pred. No. 20		
	Matches 54	Conservative 0	Mismatches 39	Indels 0	Gaps 0
QY	13	agactcctgcacgctctgtgtacagatgatgaagagaagatggtgtatgctcgtct	72		
DB	217	ACACTCTCTGTGAGACATCTCTTCCATATGGGGACACCATCTTCTGCCCTTTCTGCTCA	158		
QY	73	tgcactagagatcagcgccctctcttgaagctt	105		
DB	157	TGTTTCAGGCTCATCTCCAAAGTGTAATCAAGCTTT	125		

[illegible]

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus- 1 (bases 1 to 601)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly ,M., Rose,M., Rose,R., Stokes,R., Tinger,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0190 Row: E Column: 03
Seq primer: CGTGTAAACAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 601.
Location/Qualifiers
1. .601
/organism="Mus musculus"
/strain="C57BL/6J"


```

BASE COUNT      85 a      53 c      52 g      70 t
ORIGIN           /note="Organ: brain"

Query Match      28.8%; Score 30.2; DB 11; Length 260;
Best Local Similarity 69.5%; Pred. No. 21;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY      15  acatctgcagttctgtgtacagatgcagcagcagatgtgtatgtgtcgtctt 73
          ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      84  ACTTCTGTGTAGTTCTGTCTGTCAGATACGCAAGACCTGCTGGGTGTCTTCTTT 26

RESULT  5
H91634      331 bp      mRNA      EST      29-NOV-1995
LOCUS      y003f12.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:241679 3', mRNA sequence.
ACCESSION  H91634
VERSION    H91634.1 GI:1087212
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 331)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisan,E., Waterson
            ,R., Williamson,A., Wohlmann,P. and Wilson,R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 291
            Source: IMAGE Consortium, LNLN.
            This clone is available royalty-free through LNLN; contact the
            IMAGE Consortium (info@image.lnl.gov) for further information.
            Seq primer: Promega -21m13
            High quality sequence stop: 291.
            Location/Qualifiers
                1..331
                    /organism="Homo sapiens"
                    /db_xref="GDB:3790812"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:241679"
                    /clone_lib="Soares fetal liver spleen INFLS"
                    /sex="male"
                    /dev_stage="20 week-post conception fetus"
                    /lab_host="DH10B (ampicillin resistant)"
                    /note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)
                    with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
                    1st strand cDNA was primed with a Pac I - oligo(dT) primer
                    [5' AACTCGACAGATTAATTAAGACCTTTTCTTTTCTTTTCTTTT 3'],
                    double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Pac I and cloned into the Pac I
                    and Eco RI sites of the modified pTV73 vector. Library
                    went through one round of normalization. Library
                    constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      92 a      64 c      63 g      106 t      6 others
ORIGIN

Query Match      28.8%; Score 30.2; DB 11; Length 331;
Best Local Similarity 69.5%; Pred. No. 23;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Y      15  acatctgcagttctgtgtacagatgcagcagcagatgtgtatgtgtcgtctt 73

```

Db 235 ACTTCTGTGAGTTCTGTGTGACATGCAAGACACTGCGTGTGGGTGTGTCTTT 293

RESULT 6
AM006058
LOCUS AM006058 359 bp mRNA EST 20-OCT-2000
DEFINITION w281d05.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2565225 3', mRNA sequence.
ACCESSION AM006058
VERSION AM006058.1 GI:5854836
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 359)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskajuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lnl.gov/db/rp/image/image.html
Insert Length: 1808 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
FEATURES
source
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2565225"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 106 a 66 c 69 g 117 t 1 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 359;
Best Local Similarity 69.5%; Pred. No. 23;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 15 acatcctgcagctctcgtgtacgatgatgcagcagcagtggtatgtctgtctctt 73
|| ||||| ||||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
Db 256 ACTTCTGTGAGTTCTGTGTGACATGCAAGACACTGCGTGTGGGTGTGTCTTT 314

RESULT 7
AA252861
LOCUS AA252861 361 bp mRNA EST 06-AUG-1997
DEFINITION z776c01.s1 Soares.NHMPu.S1 Homo sapiens cDNA clone IMAGE:669312
ACCESSION AA252861
VERSION AA252861.1 GI:1885483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 361)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1134 Std Error: 0.00
Seq primer: -41m3 fwd. ET from Amersham.
Location/Qualifiers
FEATURES
source
1..361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:669312"
/clone_lib="Soares.NHMPu.S1"
/tissue_type="pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI. Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NHM, pregnant uterus NBPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 108 a 73 c 66 g 114 t
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 361;
Best Local Similarity 69.5%; Pred. No. 23;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 15 acatcctgcagctctcgtgtacgatgatgcagcagcagtggtatgtctgtctctt 73
|| ||||| ||||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
Db 183 ACTTCTGTGAGTTCTGTGTGACATGCAAGACACTGCGTGTGGGTGTGTCTTT 241

RESULT 8
BE223043
LOCUS BE223043 386 bp mRNA EST 03-JUL-2000
DEFINITION hu46d12.x1 NCI_CGAP_Brnal Homo sapiens cDNA clone IMAGE:317311 3', mRNA sequence.
ACCESSION BE223043
VERSION BE223043.1 GI:8910361
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 386)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Robert Jenkins, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D. and M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center

BASE COUNT 129 a 74 c 78 g 117 t 4 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 402;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 141 ACTTCCTGTGAGTCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 83

RESULT 11
R76040 404 bp mRNA EST 06-JUN-1995
LOCUS Y160C01.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
DEFINITION IMAGE:143616 5', mRNA sequence.
ACCESSION R76040
VERSION R76040.1 GI:850722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 712
High quality sequence stops: 293
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 712 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 293.

FEATURES
source
1.404
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:552797"
/db_xref="taxon:9606"
/clone="IMAGE:143616"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Placenta; Vector: pTZ19 (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGAGACAAATTCGCGCCGCGAGCAATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTZ19 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 128 a 72 c 82 g 120 t 2 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 11; Length 404;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Query Match 28.8%; Score 30.2; DB 10; Length 416;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 68 ACTTCCTGTGAGTCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 10

RESULT 13
A1076519 418 bp mRNA EST 29-SEP-1998
LOCUS OZ09104.x1 Soares fetal_liver_spleen_1NF15_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:1674847 3', mRNA sequence.
ACCESSION A1076519
VERSION A1076519.1 GI:3405697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

BASE COUNT 131 a 78 c 77 g 128 t 2 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 416;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 140 ACTTCCTGTGAGTCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 82

RESULT 12
AV725870 416 bp mRNA EST 16-OCT-2000
LOCUS AV725870 HTC Homo sapiens cDNA clone HTCAMD10 5', mRNA sequence.
DEFINITION AV725870
ACCESSION AV725870
VERSION AV725870.1 GI:10831692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu,
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1.416
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCAMD10"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

EGTTACCAATCTGAGTGGAGCGCCGCAATTTTTTTTTTTTTTTT 3'] ,
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH). "

BASE COUNT 134 a 91 c 82 g 114 t 1 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 11; Length 422;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 15 acatccctgcagttctgtgtacgatgatgcagcagcatgtgtatgtctctctt 73
Db 342 ACTTCCTGTGAGTTCCTGTCTGACATCTGCAAGACATCGCTGTGGTGTGTGCTTT 284

Search completed: January 17, 2002, 06:42:54
Job time: 3081 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH nucleic - nucleic search, using sw model

Run on: January 17, 2002, 08:48:57 ; Search time 2206.96 Seconds
(without alignments)
2900.327 Million cell updates/sec

Title: US-09-823-101-8

Perfect score: 388
Sequence: 1 acgcctggcgcacagtcgagc.....acaaacatacagaccacaaa 388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sy:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_inv:*
33: em_htgo_inv:*
34: em_htg_hum:*
35: em_htg_inv:*
36: em_htg_inv:*
36: em_htg_inv:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	36.6	9.4	110000	2	LMFLCHR36_27	Continuation (28 O
2	36.4	9.4	69156	2	AC090542	AC090542 Homo sapi
3	35.6	9.2	28512	2	AC013069	AC013069 Drosophi
4	35.6	9.2	311464	3	AE003431	AE003431 Drosophi
5	35.4	9.1	564	6	AX184941	AX184941 Sequence
6	35.2	9.1	153452	2	AC044895	AC044895 Homo sapi
7	35.2	9.1	202397	2	AC074394	AC074394 Homo sapi
8	35	9.0	170541	2	AC083944	AC083944 Homo sapi
9	34.2	8.8	91842	2	AC006468	AC006468 Homo sapi
10	33.8	8.7	884	2	AC074974	AC074974 Giardia i
11	33.8	8.7	146150	2	AC068333	AC068333 Homo sapi
12	33.6	8.7	73522	2	AC006173	AC006173 Homo sapi
13	33.6	8.7	84722	9	AL133420	AL133420 Human DNA
14	33.6	8.7	159967	2	AC048356	AC048356 Homo sapi
15	33.4	8.6	167	9	HUMPEPC7	M18665 Human pepsi
16	33.4	8.6	172	9	HUMPEPCA7	M23075 Human pepsi
17	33.4	8.6	1318	9	HUMPGCA	J04443 Homo sapien
18	33.4	8.6	1318	11	G28579	G28579 human SRS S
19	33.4	8.6	1418	6	AX014678	AX014678 Sequence
20	33.4	8.6	74371	9	AC005369	AC005369 Homo sapi
21	33.4	8.6	81914	9	HSJD63M2	AL121906 Human DNA
22	33.4	8.6	106711	9	AL365205	AL365205 Human DNA
23	33.4	8.6	148042	2	AC023227	AC023227 Homo sapi
24	33.4	8.6	186092	9	AC020556	AC020556 Homo sapi
25	33.2	8.6	6741	6	E32987	E32987 Gene encodi
26	33.2	8.6	112107	9	AL139320	AL139320 Human DNA
27	33.2	8.6	177489	2	AC092407	AC092407 Papio cyn
28	33	8.5	7218	6	166494	166494 Sequence 14
29	32.8	8.5	153201	9	AC006454	AC006454 Homo sapi
30	32.8	8.5	202028	2	AC093086	AC093086 Homo sapi
31	32.6	8.4	42265	9	AC007136	AC007136 Homo sapi
32	32.4	8.4	108280	2	AL355862	AL355862 Homo sapi
33	32.4	8.4	148915	9	AC005030	AC005030 Homo sapi
34	32.4	8.4	151805	2	AC013741	AC013741 Homo sapi
35	32.4	8.4	158408	2	AC013336	AC013336 Homo sapi
36	32.4	8.4	176238	2	AC036122	AC036122 Mus muscu
37	32.4	8.4	182975	2	AC079085	AC079085 Homo sapi
38	32.2	8.3	168497	2	AC092410	AC092410 Bos tauru
39	32	8.2	1096	9	AF022229	AF022229 Homo sapi
40	32	8.2	1276	9	MEPEPC	X59754 M fuscata m
41	32	8.2	15456	6	AX073600	AX073600 Sequence
42	32	8.2	15456	14	AF178654	AF178654 Bovine pa
43	32	8.2	32110	3	CEFA0612	Z77661 Caenorhabdi
44	32	8.2	50579	9	HS1086L22	AL109910 Human DNA
45	32	8.2	99367	2	AL358940	AL358940 Homo sapi

ALIGNMENTS

RESULT 1
LMFLCHR36_27
WPCOMMENT

Sequence split into 36 fragments				LOCUS LMFLCHR36 Accession AL499624	
Fragment Name	Begin	End			
LMFLCHR36_00	1	110000			
LMFLCHR36_01	100001	210000			
LMFLCHR36_02	200001	310000			
LMFLCHR36_03	300001	410000			
LMFLCHR36_04	400001	510000			
LMFLCHR36_05	500001	610000			
LMFLCHR36_06	600001	710000			
LMFLCHR36_07	700001	810000			
LMFLCHR36_08	800001	910000			
LMFLCHR36_09	900001	1010000			
LMFLCHR36_10	1000001	1110000			
LMFLCHR36_11	1100001	1210000			
LMFLCHR36_12	1200001	1310000			
LMFLCHR36_13	1300001	1410000			
LMFLCHR36_14	1400001	1510000			

LMFLCHR36_15	1500001	1610000
LMFLCHR36_16	1600001	1710000
LMFLCHR36_17	1700001	1810000
LMFLCHR36_18	1800001	1910000
LMFLCHR36_19	1900001	2010000
LMFLCHR36_20	2000001	2110000
LMFLCHR36_21	2100001	2210000
LMFLCHR36_22	2200001	2310000
LMFLCHR36_23	2300001	2410000
LMFLCHR36_24	2400001	2510000
LMFLCHR36_25	2500001	2610000
LMFLCHR36_26	2600001	2710000
LMFLCHR36_27	2700001	2810000
LMFLCHR36_28	2800001	2910000
LMFLCHR36_29	2900001	3010000
LMFLCHR36_30	3000001	3110000
LMFLCHR36_31	3100001	3210000
LMFLCHR36_32	3200001	3310000
LMFLCHR36_33	3300001	3410000
LMFLCHR36_34	3400001	3510000
LMFLCHR36_35	3500001	3529852

Query Match	9.4%	Score 36.6	DB 2	Length 110000
Best Local Similarity	31.8%	Pred. No. 1.6	Mismatches 174	Indels 0
Matches 81	Conservative 0			Gaps 0
OY 134	gctacgcggcctgtgagcaatgtcgtaccacgcacagatctgcccgcgattggaacatcg	193		
Db 55858	GCCACGTGGGCTCGGAGCCCTTCGTCGTCGCGCCCAAGAAAGAGGCGCTGCATCAGC	55917		
OY 194	cgagagcagctgagcggtacgcacagctaatgttgagatccgcgctcgtgtatcgcgaacg	253		
Db 55918	GCGNN	55977		
OY 254	tatgggggtacactcatagtgcatagctcttcgtcgtgaigtacacatggttcata	313		
Db 55978	NNNGGGAAGAACACAA	56037		
OY 314	ccctgtacacaatgtcgcgaacaatctgatactgacgcacccaattccaacaaca	373		
Db 56038	GAAAGGAAACAGAAAGAGACACAAACAAAACAAAAAGAAAAAAGCGAAGGAAAAA	56097		
OY 374	catacagaccacaacaa	388		
Db 56098	CAAGAGGAAACACAAA	56112		
RESULT 2				
AC090542				
LOCUS	AC090542			
DEFINITION	Homo sapiens chromosome 18 clone RP11-719K4 map 18, LOW-PASS			
SEQUENCE SAMPLING.	AC090542			
ACCESSION	AC090542.3			
VERSION	GI:14141773			
KEYWORDS	HTG; HTGS_PHASE0.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 69156)			
JOURNAL	Bliren,B., Linton,L., Nusbaum,C. and Lander,E.			
REFERENCE	Homo sapiens chromosome 18, clone RP11-719K4			
AUTHORS	Unpublished			
	2 (bases 1 to 69156)			
	Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,			
	Barna,N., Bastien,J., Boguslavsky,L., Boukhalter,B., Brown,A.,			
	Camatare,J., Campolano,A., Choepel,Y., Colangelo,M., Collins,S.,			
	Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,			
	Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,			
	Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,			
	Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,			

TITLE
JOURNAL
COMMENT

TITLES

JOURNAL

COMMENT

Jones, C., Karitas, A., LaRocque, K., Lamazeres, R., Landers, T.,
 Lehoczyk, J., Levine, R., Liu, G., McLean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K.,
 McPeeters, R., Meldrum, J., Menus, L., Minova, T., Menga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norby, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkharn, P., Pierre, N., Pollara, V., Raymond, C., Retts, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
 Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,
 Sounez, C., Spencer, B., Stange, T., Thomann, N., Stojanovic, N.,
 Strauss, N., Sudirmanatan, A., Talamas, J., Teste, S., Theodore, J.,
 Travers, M., Travis, N., Triggilo, J., Vassiliou, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J.,
 Zemsek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (02-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 18, 2001 this sequence version replaced g131488001.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/MK/RepeatMasker.html>
 Genome Center

```

Center:Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
Project Information
Center project name: L12330
Center clone name: 719_K_4
-----
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be generic-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
707 806: contig of 706 bp in length
807 1557: contig of 751 bp in length
1558 1657: gap of 100 bp
1658 2399: contig of 742 bp in length
2400 2499: gap of 100 bp
2500 3242: contig of 743 bp in length
3243 3342: gap of 100 bp
3343 4045: contig of 703 bp in length
4046 4145: gap of 100 bp
4146 4884: contig of 739 bp in length
4885 4984: gap of 100 bp
4985 5722: contig of 738 bp in length
5723 5822: gap of 100 bp
5823 6573: contig of 751 bp in length
6574 6673: gap of 100 bp
6674 7395: contig of 722 bp in length
7396 7495: gap of 100 bp
7496 8326: contig of 731 bp in length
8327 8326: gap of 100 bp
8327 9065: contig of 739 bp in length
9066 9165: gap of 100 bp
9166 9915: contig of 750 bp in length
9916 10015: gap of 100 bp
10016 10722: contig of 707 bp in length
10723 10822: gap of 100 bp
10823 11580: contig of 756 bp in length
11581 11680: gap of 100 bp
11681 12417: contig of 737 bp in length
12418 12517: gap of 100 bp
12518 13244: contig of 727 bp in length
13245 13344: gap of 100 bp

```


- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

FEATURES	Location/Qualifiers
SOURCE	1. 28512 /organism="Drosophila melanogaster" /db_xref="taxon:7227"
BASE COUNT	7901 a 6188 c 6339 g 8084 t
ORIGIN	

Query Match	9.28;	Score 35.6;	DB 2;	Length 28512;
Best Local Similarity	54.68;	Pred. NO. 3.4;		
Matches 71; Conservative	0;	Mismatches 59;	Indels 0;	Gaps 0;

QY	244	tacgaagaagcatttggggttacacttccttagtgatccgcctagctctcgctctagtgaaca	303
Db	7390	TACCACCCGTCGAAGGGGTAAATCATCACTGTGGTAAAGACTTTCCGATTGTGCCA	7449
QY	304	tgggtcaatacctfgrtaaacaaatttgcgaaaaaaatctcgatctagcagcatcccaattt	363
Db	7450	TCCGTAATAATCAAGAAGTACAGCTATTGGTTAAAGGCACCTTTAACAAATTCATTTCACTAG	7505

QY	364	ccaacaacaa	373
Db	7510	CTATAAATA	7519

RESULT	4
AE003431/c	
LOCUS	
DEFINITION	AE003431 311464 bp DNA INV 04-OCT-2000 Drosophila melanogaster genomic scaffold 142000013386054 section 15 of 35 complete sequence.
ACCESSION	AE003431 AE002566
VERSION	AE003431.2 GI:10728461
FEATURES	
ORIGIN	

SOURCE ORGANISM

REFERENCE
AUTHORS

AUTHORS
 Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
 Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
 George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
 Sutton, G.G., Wortman, J.R., Vandeell, M.D., Zhang, Q., Chen, L.X.,
 Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,
 Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
 Miklos, G.L., Adril, J.F., Agbayani, A., An, H.J.,
 Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A.,
 Baxendale, J.K., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
 Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borokova, D.,
 Botchan, M.R., Bouck, J., Brokstein, P., Brotlier, P., Burris, K.C.,
 Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahlke, C., Davidov, L.B., Davies, P., de
 Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
 Dodson, K., Dunn, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Dublin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,
 Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.,
 Gelibert, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Havelly, D., Heiman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wei, M.H., Ikegawa, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Krat, C.,
 Kravitz, S., Kulp, D., Lai, Z., Lasok, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Pauley, J.M., Palazolo, M., Pittman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,

TITLE	2 (bases 1 to 311464)
JOURNAL	Adams,M.D., Celinker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
MEDLINE	Direct Submission
REFERENCE	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
AUTHORS	Rockville, MD, USA
JOURNAL	On Oct 9, 2000 this sequence version replaced gl:7290491.
COMMENT	Location/Qualifiers
FEATURES	I .311464
Source	

gene

CDS

gene

CDS

[illegible]

```
KDLWEKFKLCTENVTITKSGROMFPQMKFRVSGLDAAKAYILLIDIVAADDYRYKFHN
SRMNVAKADPEMEKRMV IHPDSTPTGDOMQKVPFHKLKTNNISQKHGFEVSTTL
NSMKHYOPRHLVANDIILKPYSTFRITYKEKEFTAVRAYONEKITQOLKIDNNPRA
KGRFDTGAKREKNCYIQAOLMSNRGSDSKLNPYHVSSSRAPLHLGHAGRPHLHPHA
ALDNOQDDDDDKLLDVYRPPQSPLLPLSHSLQOMHAHQHSAALAWFNHLAAGAGAS
EHAHAANAANAEDALRRRLQADADVERSDSDSCSVSGSTGAFRPTSTGSKPA
VGAAMAAAAGLNGGSGSPSPNTSVGRPIHSPHLLPYLYPHGLYPPPHGLHJHPA
AQAASPPAGLPGILLFMAOLALAAOHAPALFGHAYAAAGHMPVSPLOGLKSREPSYL
PGLSGSAFDAVTGCSNANRSGDPPRGSGGGLGCGVVENGRPSLSSSPRRASHPPT
RPIWSPTTPPSLMOKOPRGAGAGVAAOSQHSPELSMEKMNGLVQHNHNGSAAAA
AALQAEAAQHNNHTQHNNQOQOQHQQOHPHPHQHHLHSHHGATTCG
TDQYGGCGRGSGSGSG"
complement(join(47962..48065,48215..>48506))
/mrna
/gene="CG12686"
/product="CT35553"
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
complement(<47962..>48506)
/gene="CG12686"
/map="AC2-4C2"
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
complement(join(47962..48065,48215..48233))
/gene="CG12686" gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
/protein_id="AAFA5947.1"
/db_xref="GI:7290494"
/translacion="MOTDSTLQEEFFGDIQPLTHOPASAHCISSKTSRYPLSM"
join(<126141..126277,126550..>126709)
/gene="CG12685"
/product="CT35551"
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence="not_experimental"
<126141..>126709
/gene="CG12685"
/map="AC3-4C4"
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence="not_experimental"
join(126141..126277,126550..126709)
/gene="CG12685"
/note="CG12685 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence="not_experimental"
/protein_id="AAR45948.1"
/db_xref="GI:7290495"
/translacion="MDDGAASRACPETGIMVRCRMODAGCMLAECQWKLILDLGL
SCVASPIVATOLTRTROIAPAPPTAYRPPRPAQOQHHPOMAAVGVGK"
join(140773..141143,141852..143220,143398..143553,
143862..143739,143803..143885,144067..>144128)
/gene="CG3556"
/product="CT11944"
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
<140773..>144128
/gene="CG3556"
/map="AC3-4C4"
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
join(141852..143220,143398..143593,143662..143739,
143803..143885,144067..144128)
/gene="CG3556" gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
/protein_id="AAR45949.1"
```

```
/db_xref="GI:7290496"
/translacion="MLSSSPPWILLVLEFFANGSATNVVSLAENVTLATLPSNPIS
VSEFTSSVAPRISTELEENQTAASSNLTNNNEADDEPTGNSNNTSSHSRINLPSS
NSNIDNANSSSVSLSTTTSALISGVDOQDSLLAPPTALNGSGTPEHOTIQAPP
TKGQGEALIRALDMLKERKASDYGMGNDTHVYIILKELSGRDNDSVDGVQVOTEL
EDTSLSEMEIEILAMLDHNTLPKPDIDKLARYVIALSGCDPKHFNHGLVATL
OHHBPADIDIEALTTISACSSAAHVRRKQIRRLIDISGVDQSDVAIAVYILARCI
VTDHRRHILQHFVRPARGLATTODOGSFCSLRSTALAMQALODIETVDPAGHNNRTA
ASRTILSRORDGCMSEBPDODGEPDIOGVLTLDITLALGKGLCAVRLQCHVIR
ESSDPTENGEPKLVAVPGLSSSAESDSAKNISTYVTLWVGSNVESFSLVSPKNTS"
Query Match 9.2%; Score 35.6; DB 3; Length 311464;
Best Local Similarity 54.6%; Pred. No. 3.1;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
OY 244 tacgcaaaactatggggatcacatcagtagtgcacgtcgtcgtgtagtgcacaa 303
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99101 TACGACACGTCMAAGGCTTAATATGATGATGCTTAAGGTAAGTTCGCTAATTTGTCCAA 99042
OY 304 tgggtcaatcctgtgtacacaattgtgcacaacatcgatcgtcgcgtacccaattt 363
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99041 TCGCTAAATACAAAGGTACACGTAATTTGTTAAAGGACCTTAACAAATTTACCTAG 98982
OY 364 ccaaacacaa 373
|| | | | | |
Db 98981 CTATATAATA 98972
RESULT 5
AX184941 564 bp DNA PAT 06-AUG-2001
LOCUS AX184941
DEFINITION Sequence 636 from Patent WO0142467.
ACCESSION AX184941
VERSION AX184941.1 GI:15136315
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 564)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
JOURNAL Patent: WO 0142467-A 636 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source 1..564
/jorganism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 157 a 128 c 138 g 123 t 18 others
ORIGIN
Query Match 9.1%; Score 35.4; DB 6; Length 564;
Best Local Similarity 60.7%; Pred. No. 4.4;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY 259 gggatcacatcagtagtgcacgtcgtcgtgtagtgcacaaatgggtcaatcctgg 318
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 GGGCTAATCATTTGGGNCATAGCTGTGTTCCGTGNGGTGAATAATGNTTAATNCGCC 494
OY 319 tacacaattgtcgacaacatcga 342
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 495 TTACACAANTTTCACCAACCA 518
RESULT 6
AC044895 153452 bp DNA HTG 25-JUN-2000
LOCUS AC044895
DEFINITION Homo sapiens chromosome 4 clone RP11-654J13 map 4, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION AC044895
VERSION AC044895.2 GI:8705180
```

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 153452)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chrmosome 4, clone RP11-654J13
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 153452)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G.,
Campioano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,D.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Hottom,L.,
Howland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidim,J., Menes,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teste-S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7543863.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR

Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L9618
Center clone name: 654_J13

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142959 bases at least Q40
Consensus quality: 148595 bases at least Q30
Consensus quality: 150659 bases at least Q20
Insert size: 180000; agarose-fp
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 1177 1276: gap of 100 bp
* 1277 2611: contig of 1335 bp in length
* 2612 2711: gap of 100 bp
* 2712 3339: contig of 628 bp in length
* 3340 3439: gap of 100 bp
* 3440 5357: contig of 1918 bp in length

* 5358 5457: gap of 100 bp
* 5458 7572: contig of 2115 bp in length
* 7573 7672: gap of 100 bp
* 7673 10775: contig of 3103 bp in length
* 10776 10875: gap of 100 bp
* 10876 13352: contig of 2477 bp in length
* 13353 13452: gap of 100 bp
* 13453 17169: contig of 3717 bp in length
* 17170 17269: gap of 100 bp
* 17270 20719: contig of 3450 bp in length
* 20720 20819: gap of 100 bp
* 20820 25578: contig of 4759 bp in length
* 25579 25678: gap of 100 bp
* 25679 32478: contig of 6800 bp in length
* 32479 32578: gap of 100 bp
* 32579 38908: contig of 6330 bp in length
* 38909 39008: gap of 100 bp
* 39009 48521: contig of 9513 bp in length
* 48522 48621: gap of 100 bp
* 48622 60786: contig of 12165 bp in length
* 60787 60886: gap of 100 bp
* 60887 74000: contig of 13114 bp in length
* 74001 74100: gap of 100 bp
* 74101 95823: contig of 21723 bp in length
* 95824 95923: gap of 100 bp
* 95924 153452: contig of 57529 bp in length.
Location/Qualifiers
1. 153452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-654J13"
/clone_lib="RP11 Human Male BAC"
1. 1176
/note="assembly-fragment"
1277. 2611
/note="assembly-fragment"
clone_end:SP6
vector_side:left
2712. 3339
/note="assembly-fragment"
clone_end:T7
vector_side:right
3440. 5357
/note="assembly-fragment"
5458. 7572
/note="assembly-fragment"
7673. 10775
/note="assembly-fragment"
10876. 13352
/note="assembly-fragment"
13453. 17169
/note="assembly-fragment"
17270. 20719
/note="assembly-fragment"
20820. 25578
/note="assembly-fragment"
25679. 32478
/note="assembly-fragment"
32579. 38908
/note="assembly-fragment"
39009. 48521
/note="assembly-fragment"
48622. 60786
/note="assembly-fragment"
60887. 74000
/note="assembly-fragment"
74101. 95823
/note="assembly-fragment"
95924. 153452
/note="assembly-fragment"

BASE COUNT 44037 a 29500 c 30194 g 48117 t 1604 others


```

misc_feature      clone_end:T7
                  vector_side:right
80348..92765
misc_feature      /note="assembly_name:Contig23"
92866..106152
misc_feature      /note="assembly_name:Contig24"
106253..121067
misc_feature      /note="assembly_name:Contig25"
121168..136366
misc_feature      /note="assembly_name:Contig26
                  clone_end:Sp6
                  vector_side:left"
136467..158514
misc_feature      /note="assembly_name:Contig27"
158615..180860
misc_feature      /note="assembly_name:Contig28"
180961..202397
misc_feature      /note="assembly_name:Contig29"
BASE COUNT      62861 a 38977 c 39217 g 59228 t 2114 others
ORIGIN

```

```

Query Match
Best Local Similarity 9.1%; Score 35.2; DB 2; Length 202397;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 250 aagctatggggtacactagatgcatagcgtactgctgtagtacaatggctc 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151231 AAGTTATGTTGCACATGATTTGTTGTTACATGATTAATTTCTTTAGTGGGATTTGTG 151172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 aatactgtgacacattgtgcgaacacgatactgaacgcat 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151171 GGAACCTGTGCACCCATCATCACTAGACATGATACACTGCACCAT 151128

```

```

RESULT 8
AC083944/c      AC083944 170541 bp DNA HTG 25-JUL-2001
LOCUS          Oryza sativa clone OSJNBa0047G15, WORKING DRAFT SEQUENCE, 4
DEFINITION    unordered pieces.
ACCESSION    AC083944
VERSION      AC083944.6 GI:15011679
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Oryza sativa.
ORGANISM     Oryza sativa
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Eriharoidae; Oryzaceae; Oryza.
REFERENCE    1 (bases 1 to 170541)
AUTHORS      McComble,W.R.
TITLE        Rice genomic sequence
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 170541)
AUTHORS      McComble,W.R.
TITLE        Direct Submission
JOURNAL      Submitted (07-OCT-2000) Lila Annenberg Hazen Genome Center, Cold
              Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
              NY 11724, USA
COMMENT      On Jul 25, 2001 this sequence version replaced gi:14209715.
              ----- Genome Center
              Center: Lila Annenberg Hazen Genome Center, Cold Spring Harbor
              Laboratory
              Center code: CSHL
              Web site: http://www.cshl.org/genseq
              ----- Project Information
              Center project name: OSJNBa0047G15
              Center clone name: OSJNBa0047G15

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 111874: contig of 111874 bp in length
* 111875 112342: gap of unknown length
* 112343 150296: contig of 37954 bp in length
* 150297 150763: gap of unknown length
* 150764 164567: contig of 13804 bp in length
* 164568 165034: gap of unknown length
* 165035 170541: contig of 5507 bp in length.
FEATURES
      source
      1..170541
      /organism="Oryza sativa"
      /db_xref="taxon:4530"
      /clone="OSJNBa0047G15"
BASE COUNT      47036 a 36392 c 36240 g 49470 t 1403 others
ORIGIN

```

```

Query Match
Best Local Similarity 9.0%; Score 35; DB 2; Length 170541;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 89 cggacaagaacactcgtgtcaacgcaagctctgtcgaggtagaagctacgcggacctg 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141330 CGGCGACACCCACACGCGCGCCGCAAGACGCGGGGAGACATATCCGCGCGGC 141271
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 149 agcagtgctgaccagcagcagagctgtgtcccgatgtgaacattgtgagagcagctgagc 208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141270 AGGCGGCGCCCGCCCGCAAGCGCGCGCGCGCTGTAGAAAGTGGGTACGAGAGGA 141211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 209 gtaagctgcag 219
      ||| ||| |||
Db 141210 TCGCGCTGCAG 141200

```

```

RESULT 9
AC006468        AC006468 91842 bp DNA PRI 30-APR-1999
LOCUS          Homo sapiens chromosome 17, clone hRPC.159_F.22, complete sequence.
DEFINITION    AC006468
ACCESSION    AC006468
VERSION      AC006468.9 GI:4726137
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 91842)
AUTHORS      Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baker,J., Baldwin,J., Barna,N., Beckelky,R., Benn,J., Brown,A.,
              Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
              Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
              Funke,R., Gage,D., Galagan,J., Garyna,S., Gilbert,D., Grant,G.,
              Hagos,B., Healdord,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
              Karatas,A., Lehoczeky,J., Lieu,C., Locke,R., Macdonald,P.,
              Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
              Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Naylor,J., Nijloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
              Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
              Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
              Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
JOURNAL      Submitted (29-JAN-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3 (bases 1 to 91842)
AUTHORS      Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,

```

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collimore, A.,
 Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,
 Donegan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heatford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Meidrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 30, 1999 this sequence version replaced gi:4713975.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the first 91842 base pairs of this clone are being submitted.
 The remainder overlaps accession number AC004477 (WICGR project
 L286).

FEATURES

SOURCE

location/Qualifiers
 1..91842
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hrpc.159_F_22"
 /clone_lib="Peter de Jong/ human PAC library"
 /map="17"
 /chromosome="17"
 992..1023
 /rpt_family="GA-rich"
 3700..3905
 /rpt_family="MIR"
 4745..4766
 /rpt_family="(CCA)n"
 7979..7999
 /rpt_family="AT-rich"
 complement(8016..8320)
 /rpt_family="AluYo"
 complement(8875..9021)
 /rpt_family="L2"
 complement(9131..9433)
 /rpt_family="AluSg"
 9451..9479
 /rpt_family="(CA)n"
 9621..9742
 /rpt_family="L2"
 9810..10123
 /rpt_family="AluSp"
 10208..10504
 /rpt_family="AluSx"
 11398..11488
 /rpt_family="CT-rich"
 12526..12554
 /rpt_family="(TTTG)n"
 complement(12555..12839)
 /rpt_family="AluSg"
 14049..14352
 /rpt_family="AluYo"
 15115..15826
 /rpt_family="AluSp"
 15971..16277
 /rpt_family="L2"
 18378..18403
 /rpt_family="AT-rich"
 18949..19231
 /rpt_family="AluSc"
 19232..19271
 /rpt_family="(TAA)n"

repeat_region 20778..21095
 /rpt_family="AluYo"
 complement(22229..22535)
 /rpt_family="AluY"
 repeat_region 22617..22949
 /rpt_family="AluYa8"
 complement(23016..23318)
 /rpt_family="AluSp"
 23386..23677
 /rpt_family="AluSx"
 23680..23975
 /rpt_family="AluSx"
 24059..24199
 /rpt_family="MIR"
 24231..24275
 /rpt_family="MIR"
 24243..24282
 /rpt_family="L2"
 complement(25339..25639)
 /rpt_family="AluSx"
 25797..26082
 /rpt_family="AluSg"
 26083..26108
 /rpt_family="(CAT)n"
 complement(26133..26437)
 /rpt_family="AluYb8"
 27293..27319
 /rpt_family="(T)n"
 28155..28189
 /rpt_family="(TG)n"
 32424..32445
 /rpt_family="AT-rich"
 34286..34579
 /rpt_family="AluSg"
 complement(34706..35003)
 /rpt_family="AluSg"
 37726..38024
 /rpt_family="AluYb"
 38112..38412
 /rpt_family="AluYb"
 38719..38746
 /rpt_family="(TTTGG)n"
 complement(38747..39033)
 /rpt_family="AluSg"
 complement(39034..39380)
 /rpt_family="L2"
 complement(39488..39537)
 /rpt_family="L2"
 39538..39839
 /rpt_family="AluSg1"
 complement(39840..40127)
 /rpt_family="L2"
 40128..40447
 /rpt_family="AluSx"
 complement(40448..41238)
 /rpt_family="L2"
 41734..42027
 /rpt_family="AluSg"
 42036..42319
 /rpt_family="AluSg"
 42403..42544
 /rpt_family="AluSx"
 42545..42843
 /rpt_family="AluSx"
 42844..43008
 /rpt_family="AluSx"
 complement(43203..43409)
 /rpt_family="MIR"
 43385..43515
 /rpt_family="L2"
 complement(44174..44245)
 /rpt_family="MLT1J"
 44246..44538

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: L10071
Center clone name: 301-L-7

----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 131662 bases at least Q40
Consensus quality: 138374 bases at least Q30
Consensus quality: 141535 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 143750; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1408: contig of 1408 bp in length
* 1409 1508: gap of 100 bp
* 1509 3382: contig of 1874 bp in length
* 3383 3482: gap of 100 bp
* 3483 6192: contig of 2710 bp in length
* 6193 6292: gap of 100 bp
* 6293 9254: contig of 2962 bp in length
* 9255 9354: gap of 100 bp
* 9355 11843: contig of 2489 bp in length
* 11844 11943: gap of 100 bp
* 11944 13616: contig of 1673 bp in length
* 13617 13716: gap of 100 bp
* 13717 17351: contig of 3635 bp in length
* 17352 17451: gap of 100 bp
* 17452 20793: contig of 3342 bp in length
* 20794 20893: gap of 100 bp
* 20894 24809: contig of 3916 bp in length
* 24810 24909: gap of 100 bp
* 24910 28434: contig of 3525 bp in length
* 28435 28534: gap of 100 bp
* 28535 32973: contig of 4439 bp in length
* 32974 33073: gap of 100 bp
* 33074 38534: contig of 5461 bp in length
* 38535 38634: gap of 100 bp
* 38635 43314: contig of 4680 bp in length
* 43315 43414: gap of 100 bp
* 43415 49541: contig of 6127 bp in length
* 49542 49641: gap of 100 bp
* 49642 56631: contig of 6990 bp in length
* 56632 56731: gap of 100 bp
* 56732 61161: contig of 4430 bp in length
* 61162 61261: gap of 100 bp
* 61262 67595: contig of 6334 bp in length
* 67596 67695: gap of 100 bp
* 67696 73214: contig of 5519 bp in length
* 73215 73314: gap of 100 bp
* 73315 81193: contig of 7878 bp in length
* 81193 81292: gap of 100 bp
* 81293 89216: contig of 7924 bp in length
* 89217 89316: gap of 100 bp
* 89317 97756: contig of 8440 bp in length
* 97757 97856: gap of 100 bp

FEATURES
Source

* 97857 109268: contig of 11412 bp in length
* 109269 109368: gap of 100 bp
* 109369 121621: contig of 12253 bp in length
* 121622 121721: gap of 100 bp
* 121722 133112: contig of 11391 bp in length
* 133113 133212: gap of 100 bp
* 133213 146150: contig of 12938 bp in length.

Location/Qualifiers

1. 146150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-301L7"
/clone_lib="RPC1-11 Human Male BAC"
1. 1408
/note="assembly-fragment"
1509. 3382
/note="assembly-fragment"
3483. 6192
/note="assembly-fragment"
6293. 9254
/note="assembly-fragment"
9355. 11843
/note="assembly-fragment"
11944. 13616
/note="assembly-fragment"
clone_end:56
vector_side:right
13717. 17351
/note="assembly-fragment"
17452. 20793
/note="assembly-fragment"
20894. 24809
/note="assembly-fragment"
24910. 28434
/note="assembly-fragment"
28535. 32973
/note="assembly-fragment"
33074. 38534
/note="assembly-fragment"
38635. 43314
/note="assembly-fragment"
43415. 49541
/note="assembly-fragment"
49642. 56631
/note="assembly-fragment"
56732. 61161
/note="assembly-fragment"
61262. 67595
/note="assembly-fragment"
67696. 73214
/note="assembly-fragment"
clone_end:77
vector_side:right
73315. 81192
/note="assembly-fragment"
81293. 89216
/note="assembly-fragment"
89317. 97756
/note="assembly-fragment"
97857. 109268
/note="assembly-fragment"
109369. 121621
/note="assembly-fragment"
121722. 133112
/note="assembly-fragment"
133213. 146150
/note="assembly-fragment"

BASE COUNT 39106 a 32795 c 32929 g 38913 t 2407 others
ORIGIN

Db 71667 CTCTCCGTCTTACCTTTGACCTGACCTGACCAACATCATATTAACATACAAAC 71726

QY 347 gacgcataccaataattccaaataacataacgaccaa 386
 11 111111 11 111 111
 Db 71727 AACATATTCACCAAGTCTTAAGGAAATTTGCTCTCAA 71766

RESULT 14
 AC048356/c AC048356 159967 bp DNA HTG 06-MAY-2000
 DEFINITION Homo sapiens chromosome 20 clone RP11-250P20 map 20, WORKING DRAFT
 AC048356 SEQUENCE 18 unordered pieces.
 AC048356.2 GI:7712163
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 159967)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abram, H., Allen, N.,
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 JOURNAL Homo sapiens chromosome 20, clone RP11-250P20
 REFERENCE 2 (bases 1 to 159967)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abram, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,
 Campiolo, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margus, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
 Melidim, J., Menus, L., Minova, T., Miranda, C., Mlenka, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 6, 2000 this sequence version replaced gl:7549675.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L5951
 Center clone name: 250_P-20

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151130 bases at least Q40
 Consensus quality: 155479 bases at least Q30
 Consensus quality: 157050 bases at least Q20
 Insert size: 160000; agarose-1p
 Insert size: 158267; sum-of-contents
 Quality coverage: 4.6 in Q20 bases; sum-of-contents
 Quality coverage: 4.6 in Q20 bases; sum-of-contents

* NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1473: contig of 1473 bp in length
 * 1474 1573: gap of 100 bp
 * 1574 3513: contig of 1940 bp in length
 * 3514 3613: gap of 100 bp
 * 3614 6910: contig of 3297 bp in length
 * 6911 7010: gap of 100 bp
 * 7011 10072: contig of 3062 bp in length
 * 10073 10172: gap of 100 bp
 * 10173 14497: contig of 4325 bp in length
 * 14498 14597: gap of 100 bp
 * 14598 19734: contig of 5137 bp in length
 * 19735 19834: gap of 100 bp
 * 19835 26423: contig of 6589 bp in length
 * 26424 26523: gap of 100 bp
 * 26524 33395: contig of 6872 bp in length
 * 33396 33495: gap of 100 bp
 * 33496 41230: contig of 7735 bp in length
 * 41231 41330: gap of 100 bp
 * 41331 48383: contig of 7053 bp in length
 * 48384 48483: gap of 100 bp
 * 48484 57555: contig of 9072 bp in length
 * 57556 57655: gap of 100 bp
 * 57656 69694: contig of 12039 bp in length
 * 69695 69794: gap of 100 bp
 * 69795 81173: contig of 11379 bp in length
 * 81174 81273: gap of 100 bp
 * 81274 93212: contig of 11939 bp in length
 * 93213 93312: gap of 100 bp
 * 93313 106989: contig of 13677 bp in length
 * 106990 107089: gap of 100 bp
 * 107090 122757: contig of 15668 bp in length
 * 122758 122857: gap of 100 bp
 * 122858 140836: contig of 17979 bp in length
 * 140837 140936: gap of 100 bp
 * 140937 159967: contig of 19031 bp in length.

FEATURES
 source
 1. 159967
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="20"
 /clone="RP11-250P20"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1473
 /note="assembly_fragment"
 misc_feature
 1574. 3513
 /note="assembly_fragment"
 misc_feature
 3614. 6910
 /note="assembly_fragment"
 misc_feature
 7011. 10072
 /note="assembly_fragment"
 misc_feature
 10173. 14497
 /note="assembly_fragment"
 misc_feature
 14598. 19734
 /note="assembly_fragment"
 misc_feature
 19835. 26423
 /note="assembly_fragment"
 misc_feature
 26524. 33395
 /note="assembly_fragment"
 misc_feature
 33496. 41230
 /note="assembly_fragment"
 misc_feature
 41331. 48383
 /note="assembly_fragment"
 misc_feature
 48384. 57555
 /note="assembly_fragment"
 misc_feature
 57556. 69694
 /note="assembly_fragment"
 misc_feature
 69695. 81173
 /note="assembly_fragment"
 misc_feature
 81174. 93212
 /note="assembly_fragment"
 misc_feature
 93213. 106989
 /note="assembly_fragment"
 misc_feature
 106990. 122757
 /note="assembly_fragment"
 misc_feature
 122758. 140836
 /note="assembly_fragment"
 misc_feature
 140837. 140936
 /note="assembly_fragment"
 misc_feature
 140937. 159967
 /note="assembly_fragment"

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:46:54 ; Search time 218.99 Seconds
(without alignments)
1518.983 Million cell updates/sec

Title: US-09-823-101-8
Perfect score: 388
Sequence: 1 actgcctggcgccacagctgacg.....acaacatacagcaccacaa 388

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	10.0	741	AA280586	Human colon cancer
2	35.4	9.1	564	AAH69362	Human cervical can
3	33.4	8.6	819	AAK39801	Gastric cancer cell
4	33.4	8.6	1357	AAH57422	Human stomach cell
5	33.4	8.6	1358	AAK39788	Gastric cancer ass
6	33.4	8.6	1418	AAZ77523	Human ovarian tumo
7	33.2	8.6	6741	AAA10595	Gene encoding a su
8	33	8.5	1397	AAZ56732	Human transmembran
9	32	8.2	15456	AAK85008	Nucleotide sequenc
10	31.8	8.2	1366	AAAD09486	Human gastricsin c
11	31.4	8.1	1255	AAAD00028	3' end of human zf

12	31.4	8.1	1434	21	AAA99907	CDNA encoding huma
13	31.4	8.1	1434	21	AAC58623	Human PRO205 prote
14	31.4	8.1	3192	21	AAAD00023	Follistatin-relate
15	31.2	8.0	96109	22	AAF28548	Genomic fragment #
16	31	8.0	10732	21	AAA10594	Gene encoding a su
17	30.8	7.9	827	21	AAZ97295	Human prostate can
18	30.8	7.9	2949	21	AAAD00024	Degenerate DNA enc
19	30.8	7.9	10930	16	AAO81226	Plasmid pM16-1. C
20	30.8	7.9	10950	16	AAO81225	Plasmid pM16. Chi
21	30.6	7.9	801	19	AAV29367	Calcium ion channe
22	30.4	7.8	977	19	AAV31204	E. coli J96 pathog
23	29.8	7.7	3266	22	AAH72600	Human cervical can
24	29.8	7.7	3266	22	AAH72698	Human cervical can
25	29.8	7.7	3266	22	AAH72723	Human cervical can
26	29.6	7.6	2617	13	AAO29753	DNA sequence encod
27	29.6	7.6	3384	19	AAV11346	DNA encoding C-bet
28	29.6	7.6	3384	21	AAV12454	DNA encoding C-bet
29	29.6	7.6	3387	19	AAV11347	DNA encoding C-bet
30	29.6	7.6	3387	19	AAV12455	DNA encoding C-bet
31	29.6	7.6	3405	19	AAV11345	DNA encoding C-bet
32	29.6	7.6	3405	21	AAA12453	DNA encoding a C-b
33	29.6	7.6	3492	19	AAV11348	DNA encoding C-bet
34	29.6	7.6	3492	21	AAV12456	DNA encoding a C-b
35	29.6	7.6	3730	17	AAO31195	Group B streptococ
36	29.6	7.6	4200	17	AAT03190	Group B streptococ
37	29.6	7.6	4200	19	AAV11344	Group B streptococ
38	29.6	7.6	4200	21	AAA12452	DNA encoding wild
39	29.6	7.6	21591	20	AAV13047	Enterococcus faeca
40	29.4	7.6	1122	22	AAK84392	Mouse A236 polypep
41	29.4	7.6	1866	11	AAO04292	Encodes polypeptid
42	29.4	7.6	1949	22	AAC84391	Mouse A236 polypep
43	29.4	7.6	1949	22	AAC84416	Mouse A236 variant
44	29.4	7.6	1949	22	AAC84417	Mouse A236 variant
45	29.4	7.6	1949	22	AAC84418	Mouse A236 variant

ALIGNMENTS

RESULT 1	
AA280586	AA280586 standard; cDNA; 741 BP.
ID	AA280586;
AC	AA280586;
XX	
DT	07-APR-2000 (first entry)
XX	
DE	Human colon cancer cell line SW480 cDNA clone seq ID NO:670.
XX	
KM	Human; gene expression product; diagnosis; tumour; colon cancer;
KM	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KM	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	
OS	Homo sapiens.
XX	
PN	W09964576-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-1B01062.
XX	
PR	10-JUN-1998; 98US-0088801.
XX	
PA	(FARF) BAYER CORP.
XX	
PI	Endege WO, Steilmann KE, Astle JH, Burgess CC, Bushnell SE;
XX	Carroll F, Catino JV, Dertli A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	
DR	WPI: 2000-087220/07.
XX	
PT	Novel nucleic acids, used to develop products for the diagnosis and

PT	treatment of disorders involving unwanted cell proliferation,
PT	particularly cancers, especially colon cancer
XX	
PS	Claim 15; Page 397; 469pp; English.
XX	
CC	AA279917 to AA280766 represent double stranded cDNA clones isolated from
CC	the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC	cDNA clones can be used to generate antisense oligonucleotides which
CC	can be used for antisense therapy. Methods and products from the present
CC	invention can be used for identifying and/or classifying cancerous cells
CC	present in a human tumour, particularly in solid tumours, e.g.
CC	carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC	can be used for developing agents for the diagnosis and treatment of
CC	disorders involving unwanted cell proliferation, such as neoplasia,
CC	dysplasia or hyperplasia.
XX	
SQ	Sequence 741 BP; 183 A; 184 C; 182 G; 166 T; 26 other;
Query Match	10.0%; Score 38.8; DB 21; Length 741;
Best Local Similarity	58.1%; Pred. No. 0.0038;
Matches 61; Conservative	0; Mismatches 44; Indels 0; Gaps
Dy	223 atcgaggatccgaatctgctgcgtacgcaaaagctatgggggtacactcatagtgcatacgc 282 531 aagtggaatccnaaccttcggnaccacaagaacttgcctgaatcacatgccataagcgcataagc 590
Oy	283 tagcttcgcctgtagtgacaattggtgccaataacccgtgacacaaatt 327 Db 591 ttggtccctcgtggggggaataattgtaatacncggttacacnaatt 635
RESULT 2	
AAH69362	
ID	AAH69362 standard; CDNA; 564 BP.
AC	
XX	AAH69362;
XX	
DT	19-SEP-2001 (first entry)
XX	
DE	Human cervical cancer marker nucleic acid 636.
XX	
KW	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200142467-A2.
PD	
FD	14-JUN-2001.
XX	
Pf	08-DEC-2000; 2000WO-US33312.
XX	
PR	08-DEC-1999; 99US-0169681.
PR	21-DEC-1999; 99US-0171350.
PR	14-MAR-2000; 2000US-0189315.
PR	12-MAY-2000; 2000US-0203791.
PR	09-JUN-2000; 2000US-0210600.
PR	21-JUL-2000; 2000US-0220114.
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
PI	Schlegel R, Deeds J, Berger A, Zhao X;
XX	
DR	WPI; 2001-375006/39.
XX	
PT	New isolated nucleic acid for diagnosing and treating cervical cancer
XX	and for assessing and detecting compounds for treating the cancer -
XX	
PS	Claim 1; Page 214; 1051pp; English.
XX	
CC	The invention relates to novel genes (AAH68727-AAH73383) associated with
CC	cervical cancer with cytostatic activity. The nucleic acids and encoded
CC	polypeptides are useful: to assess if a patient is afflicted with

```

CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
S0 Sequence 564 BP; 157 A; 128 C; 138 G; 123 T; 18 other;

Query Match          9.1%; Score 35.4; DB 22; Length 564;
Best Local Similarity 60.7%; Pred. No. 0.048;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Oy 259 gggtaactcataaggtgcatagcgtgcttcgtctgtagtgacaatggtaacactg 318
    ||| | |||| | | |||| | |||| | |||| | |||| | |||| | |||| |
Db 435 gggcatactcattggncatagctgttcttccttggtgaaatgntaatnccgc 494
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 319 tacacaattgtcgcacaacatcga 342
    | |||| | || | || | || | || |
Db 495 ttcaacaatttcaccaccaaca 518
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

RESULT 3
AAK39801
ID AAK39801 standard; DNA; 819 BP.
XX
AC AAK39801;
XX
DT 02-JUL-1999 (first entry)
XX
DE Gastric cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN MO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061589.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDWIG- ) INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 553; 787pp; English.

```


[illegible]

CC	prognosis or monitoring of treatments and for investigating a
CC	predisposition to a disorder where the gene is associated with a
CC	cancer, immunopathology or neuropathology.
XX	
SO	Sequence 1357 BP; 255 A; 415 C; 357 G; 330 T; 0 other;
	Query Match 8.6%; Score 33.4; DB 22; Length 1357;
	Best Local Similarity 78.3%; Pred. No. 0.34;
	Matches 65; Conservative 0; Mismatches 16; Indels 2; Gaps 2
QY	1 actgctggcagcagatgagcagagagtgctgtgtcatcatgactgaacctac 60
DB	912 ACTGCTGGGGGCACAGTGCAGACAGAGAG-GTCCCTGTCTCCACGATGCGCTGGCA-ACCCTC 855
QY	61 agaataccagagcgagggccatgc 83
DB	854 AGAACACGACGCGAGGAGCCTGC 832
RESULT	5
AAK39788/C	
ID	AAK39788 standard; DNA; 1358 BP.
XX	
AC	AAK39788;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Gastric cancer associated gene.
XX	
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-US14679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDWIG INST CANCER RES.
PI	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	Pfeundtschuh M, Sahlin U, Scanlan MJ, Stockert E;
PI	Tureci O;
XX	
DR	WPI: 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 548; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.

Db	3642	taccaccaaccatcaacatacaaca	3659
RESULT	10		
ID	AAD09486/c		
XX	AAD09486	standard; cDNA; 1366 BP.	
AC	AAD09486;		
XX			
DT	10-SEP-2001	(first entry)	
XX			
DE	Human gastricsin cDNA from clone 1658285.		
XX			
KM	Human: aspartyl protease; ASP; Alzheimer's disease; gene therapy;		
KM	amnesia; anxiety; antidiabetic; neuroleptic; cancer; antiinfertility;		
KM	metabolic; impotence; haemostatic; endometriosis; autoimmune disorder;		
KM	Huntington's disease; Down syndrome; cerebral palsy; hepatitis; AIDS;		
KM	seasonal affective disorder; SAD; ovulatory defect; Wilson's disease;		
KM	cirrhosis; acquired immune deficiency syndrome; transgenic animal;		
KM	epilepsy; emesis; gastricsin; ss.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	location/Qualifiers	
FT	CDS	51..1217	
FT		/*tag= a	
FT		/product= "Human gastricsin"	
XX			
PN	MO200146398-A2.		
XX			
PD	28-JUN-2001.		
XX			
XX	12-DEC-2000; 2000MO-US33743.		
PF			
PR	22-DEC-1999; 99US-0470954.		
XX			
PA	(INCY-) INCYTE GENOMICS INC.		
P1	Kaser MR, Cocks BG, Loring J, Tang YT, Yue H;		
XX	WPI: 2001-418060/44.		
DR	P-PSDB; AAED4798.		
PT	Novel mammalian aspartyl proteases useful for characterizing,		
PT	diagnosing, treating, preventing Alzheimer's disease and down		
XX	associated with altered expression of the aspartyl protease		
PS	Disclosure: Page 97-98; 102pp; English.		
XX			
CC	The present sequence is human gastricsin cDNA. Gastricsin shows		
CC	structural and chemical similarity with human aspartyl protease (ASP).		
CC	The invention relates to mammalian aspartyl proteases and nucleic		
CC	molecules encoding them. The pharmaceutical composition comprising		
CC	ASP is useful for treating diseases such as Alzheimer's disease and		
CC	Down's syndrome which are associated with altered expression of ASP.		
CC	ASP cDNA is useful for producing mammalian model system. ASP sequences		
CC	are useful for characterisation, diagnosis, prevention and treatment		
CC	of conditions such as epilepsy, Alzheimer's disease, Pick's disease,		
CC	Huntington's disease, dementia, retinitis pigmentosa, multiple		
CC	sclerosis, bacterial and viral meningitis, Creutzfeldt-Jakob disease,		
CC	fatal familial insomnia, Down syndrome, cerebral palsy, myasthenia		
CC	gravis, anxiety, seasonal affective disorder (SAD), akathisia,		
CC	amnesia, diabetic neuropathy, Tourette's disorder, disorder of		
CC	prolactin production, ovulatory defects, endometriosis, disruption		
CC	of the oestrus cycle, disruption of the menstrual cycle, uterine		
CC	fibroid, autoimmune disorders, ectopic pregnancy, cancer of the		
CC	breast, galactorrhoea, disruption of spermatogenesis, abnormal sperm		
CC	physiology, cancer of the testis and prostate, prostatic, nausea,		
CC	gastritis, impotence, dysphagia, indigestion, gastric carcinoma,		
CC	anorexia, emesis, gastroparesis, pyrosis, cholecystitis, cirrhosis,		
CC	Crohn's disease, whipple's disease, gastrointestinal haemorrhage,		
CC	acquired immune deficiency syndrome (AIDS), hepatitis, jaundice,		

CC	Wilson's disease, intrahepatic cholestasis of pregnancy, adenomas
CC	heptic tumours, and carcinomas. ASP cDNA sequences are used in
CC	hybridisation and amplification technologies to distinguish similar
CC	molecules in a sample. The molecules may be used to mimic human
CC	conditions, diseases, or disorders, to produce transgenic animal
CC	models for these conditions, to monitor animal toxicology studies,
CC	clinical trials, and patient treatment profiles. ASP hybridisation
CC	probes are useful in mapping naturally occurring genomic sequence.
CC	ASP cDNA is also useful in gene therapy.
XX	
SQ	Sequence 1366 BP; 267 A; 415 C; 357 G; 327 T; 0 other;
Query Match	8.2%; Score 31.6; DB 22; Length 1366;
Best Local Similarity	77.1%; Pred. No. 1.2;
Matches 64; Conservative	0; Mismatches 17; Indels 2; Gaps 2;
Oy	1 actgtgtagccacagtgagcagaagagtgtgccttgtgcatcatgaactgaacctac 60
Db	912 ACTGCTGGGGCGACATWAGCAGAGAG-CTGCCCTGTGCACAGATGGCTGGCA-ACCCTC 855
Oy	61 agaataccaggcgagggccatgc 83
Db	854 AGAACAACCAACGCCGAGGCGCTTGGC 832
RESULT 11	
AAD00028	
ID	AAD00028 standard; cDNA; 1255 BP.
XX	
AC	AAD00028:
XX	
DT	24-JUL-2000 (first entry)
XX	
DE	3' end of human zfstaz cDNA.
XX	
KM	Follistatin-related protein; zfstaz; Transforming growth factor-beta;
KM	TGF-beta family; acute inflammatory condition; bursitis; myositis;
RW	chronic inflammatory demyelinating polyneuropathy; contact dermatitis;
KW	contact vulvovaginitis; sepsis; ulcerative colitis; dermatological;
KW	antiinflammatory; antitumor; antibacterial; immunosuppressive; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200022126-A1.
XX	
PD	20-APR-2000.
XX	
PF	05-OCT-1999; 99WO-US23179.
XX	
PR	15-OCT-1998; 98US-0173044.
XX	
PA	(ZYMO) ZYMOGENETICS INC.
XX	
P1	Conklin DC, Ellsworth JL;
P1	WPI; 2000-317983/27.
DR	
XX	
PT	New polypeptide comprising follistatin homology domain is useful for
PT	treating acute inflammatory conditions, bursitis, contact dermatitis,
PT	contact vulvovaginitis, myositis, sepsis and ulcerative colitis -
XX	
PS	Example 1; Page 119; 125pp; English.
XX	
CC	The present sequence is the 3'end of human follistatin-related protein
CC	zfstaz cDNA. This was derived on extension of EST (expressed sequence
CC	tag) sequence related to human zftsa. This EST lacked complete 5' and 3'
CC	regions. Extension reaction was carried using foetal brain, brain, spinal
CC	cord and retina Marathonm cDNA libraries as templates. zfstaz gene and
CC	its protein are useful for binding to members of the Transforming growth
CC	factor (TGF)-beta family and mediating central nervous system,
CC	reproductive, haematopoietic and bone-related activities. This is useful
CC	for treating acute inflammatory conditions, bursitis, chronic

PR	12-MAR-1999:	99US-0123957.
PR	23-MAR-1999:	99US-0125775.
PR	12-APR-1999:	99US-0128849.
PR	20-APR-1999:	99MO-US08615.
PR	28-APR-1999:	99US-013445.
PR	04-MAY-1999:	99US-0132371.
PR	14-MAY-1999:	99US-0134287.
PR	02-JUN-1999:	99MO-US12252.
PR	23-JUN-1999:	99US-0141037.
PR	20-JUL-1999:	99US-0144758.
PR	26-JUL-1999:	99US-0145698.
PR	28-JUL-1999:	99US-0146222.
PR	01-SEP-1999:	99MO-US20111.
PR	08-SEP-1999:	99MO-US20594.
PR	13-SEP-1999:	99MO-US20944.
PR	15-SEP-1999:	99MO-US21090.
PR	15-SEP-1999:	99MO-US21547.
PR	05-OCT-1999:	99MO-US23089.
PR	29-OCT-1999:	99US-0162506.
PR	29-NOV-1999:	99MO-US28214.
PR	30-NOV-1999:	99MO-US28313.
PR	30-NOV-1999:	99MO-US28409.
PR	01-DEC-1999:	99MO-US28301.
PR	01-DEC-1999:	99MO-US28634.
PR	02-DEC-1999:	99MO-US28551.
PR	02-DEC-1999:	99MO-US28564.
PR	02-DEC-1999:	99MO-US28565.
PR	16-DEC-1999:	99MO-US30095.
PR	20-DEC-1999:	99MO-US30999.
PR	30-DEC-1999:	99MO-US31274.
PR	05-JAN-2000:	2000MO-US00219.
PR	06-JAN-2000:	2000MO-US00277.
PR	06-JAN-2000:	2000MO-US00376.
PR	11-FEB-2000:	2000MO-US03565.
PR	18-FEB-2000:	2000MO-US04341.
PR	18-FEB-2000:	2000MO-US04342.
PR	22-FEB-2000:	2000MO-US04414.

PI Asakhefi AJ, Baker KP, Goddard A, Gurney AL, Herbert C, Henzel W,
PI Ashkeoff RC, Lu Y, Pan Y, Pennica D, Shelton DL, Smith V,
PI Stewart JA, Tamas D, Watanabe CK, Wood WI, Yan M;
XX WPI; 2000-572271/53.
DR P-PSDB; AAB33458.
XX
XX
PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of
PT immune related disorders, e.g. systemic lupus erythematosus, Rheumatoid
PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -
XX
XX Claim 23: Fig 89; 309pp; English
XX

The present invention describes sixty four human PRO proteins which can be used in the treatment of immune related diseases. The human PRO proteins, anti-PRO antibodies, agonists and antagonists are useful for treating and diagnosing immune related disorders. The disorders are selected from systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal disease, demyelinating diseases of the central and peripheral nervous systems, hepatobiliary diseases, inflammatory bowel disease, gluten-sensitive enteropathy and Whipple's disease, autoimmune or immune-mediated skin diseases, allergic diseases, immunological diseases of the lung, and transplantation associated diseases including graft rejection and graft-versus-host-disease. AACS8397 to AACS8578 represent PCR primers and hybridisation probes in the isolation of human PRO sequences. AACS8579 to AACS8642 and AAB33414 to AAB33477 represent human PRO polynucleotide and protein sequences given in the exemplification of the present invention.

SQ Sequence 1434 BP; 436 A; 271 C; 327 G; 400 T; 0 other,

Query Match	8.1%	Score 31.4	DB 21	Length 1434
Best Local Similarity	53.7%	Pred. No. 1.7		
Matches 65	Conservative 0	Mismatches 56	Indels 0	Gaps 0

[illegible]

QY	388	a	388
Db	1273	a	1273

RESULT 14

ID	standard; CDNA; 3192 BP
AAD00023	

AC AAD00023;

DT 24-JUL-2000 (first entry)

Follistatin-related protein zfst_a cDNA.

KM Follistatin-related protein; zfs12; Transforming growth factor-beta, beta-1
KM TGF-beta family; acute inflammatory condition; burditts; myositis;
KM chronic inflammatory demyelinating polyneuropathy; contact dermatitis
KM contact vulvovaginitis; sepsis; ulcerative colitis; dermatological;
KM antinflammatory; antitumor; antibacterial; immunosuppressive; ss.

OS Homo sapiens

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

ET

FT sig_peptide

XX

PN W0200022126-A1.

PD 20-APR-2000.

PF 05-OCT-1999; 99WO-US23179.

PR 15-OCT-1998; 98US-0173044.

PA (ZYMO) ZYMOGENETICS INC.

PI Conklin DC, Ellsworth JL;

DR WPI; 2000-317983/27.

DR P-PSDB; AAY70775.

PT New polypeptide comprising follistatin homology domain is useful for

PT contact vulvovaginitis, myositis, sepsis and ulcerative colitis -

PS Claim 31; Page 108-114; 125pp; English.

CC The present cDNA sequence encodes follistatin-related protein zfstaz2.

CC its protein are useful for binding to members of the Transforming growth

reproductive, haematopoietic and bone-related activities. This is useful

CC Inflammatory demyelinating polyneuropathy, various forms of contact

7
.
.
.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 05:58:26 : Search time 93.06 Seconds
(without alignments)
944.267 Million cell updates/sec

Title: US-09-823-101-8

Perfect score: 388

Sequence: 1 actgcctggccacacgtgacg.....acaacataacgaacacaaa 388

Scoring table: IDENTITY_NUC

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCRNUS.COMB.seq: *
6: /cgn2_6/ptodata/2/1na/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.8	10.0	741	4	Sequence 670, App
2	33	8.5	7218	1	Sequence 14, Appl
3	30.6	7.9	289	4	Sequence 17, Appl
4	30.6	7.9	289	4	Sequence 17, Appl
5	29.6	7.6	932	1	Sequence 7, Appl
6	29.6	7.6	932	1	Sequence 7, Appl
7	29.6	7.6	932	1	Sequence 7, Appl
8	29.6	7.6	2617	1	Sequence 1, Appl
9	29.6	7.6	2617	1	Sequence 1, Appl
10	29.6	7.6	2617	3	Sequence 1, Appl
11	29.6	7.6	3294	4	Sequence 1, Appl
12	29.6	7.6	3312	4	Sequence 1, Appl
13	29.6	7.6	3384	4	Sequence 1, Appl
14	29.6	7.6	3492	4	Sequence 1, Appl
15	29.6	7.6	3730	1	Sequence 1, Appl
16	29.6	7.6	3730	1	Sequence 1, Appl
17	29.6	7.6	3730	5	Sequence 1, Appl
18	29.6	7.6	4200	1	Sequence 1, Appl
19	29.6	7.6	4200	1	Sequence 1, Appl
20	29.6	7.6	4200	5	Sequence 1, Appl
21	29.6	7.6	4200	5	Sequence 1, Appl
22	28.8	7.4	7326	1	Sequence 1, Appl
23	28.8	7.4	7344	3	Sequence 1, Appl
24	28	7.2	431	3	Sequence 1, Appl
25	28	7.2	431	4	Sequence 1, Appl
26	28	7.2	972	1	Sequence 1, Appl
27	28	7.2	972	1	Sequence 1, Appl

28	27.6	7.1	540	1	Sequence 17, Appl
29	27.6	7.1	540	5	Sequence 17, Appl
30	27.6	7.1	2260	2	Sequence 1, Appl
31	27	7.0	4695	2	Sequence 1, Appl
32	27	7.0	4695	2	Sequence 57, Appl
33	27	7.0	4695	3	Sequence 57, Appl
34	27	7.0	6122	1	Sequence 1, Appl
35	27	7.0	6122	4	Sequence 1, Appl
36	26.8	6.9	30001	1	Sequence 1, Appl
37	26.8	6.9	30001	2	Sequence 1, Appl
38	26.6	6.9	11284	4	Sequence 1, Appl
39	26.4	6.8	2010	4	Sequence 5, Appl
40	26.4	6.8	2498	4	Sequence 1, Appl
41	26.2	6.8	2749	6	Sequence 1, Appl
42	26.2	6.8	4518	3	Sequence 2, Appl
43	26.2	6.8	12839	3	Sequence 1, Appl
44	26.2	6.8	33529	4	Sequence 3, Appl
45	26.2	6.8	56516	2	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-670
Sequence 670, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Stelmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Carroll, Theodore J.
APPLICANT: Derit, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 670
LENGTH: 741
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(741)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-670

Query Match 10.0%: Score 38.8; DB 4; Length 741;
Best Local Similarity 58.1%: Pred. No. 0.00046;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Oy 223 atctgagtcgcagtcgtgacgaacgcatgggggtacacatagtcacatgc 282
Db 531 aagtcgaatccnaaccttcggnacnaacgcttggcgttaacatcattgggcataagc 590
Oy 283 tagcttcgctgttagtgaacatgggtlcaatcattggtacacatt 327
Db 591 ttgtctccctggggggggaattgtgtaatnccggttcacnaattt 635

```
RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEITLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ19-pT15
US-08-232-463-14
```

```
Query Match      8.5%; Score 33; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.15;
Matches 6; Conservative 208; Mismatches 163; Indels 0; Gaps 0;

QY 12 acagtgagcagagtgctgtgcatcatgagctgaacacctacagaalaccag 71
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1436 ACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1377

QY 72 gcgaggcgcatgcgcgcgcaagaactcgtctacagcaagctgtcgcagtag 131
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1376 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1317

QY 132 aagctacgcggcctgagcagctgctacacagcagcagctgcccgcgagtaaacat 191
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1316 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1257

QY 192 tgcgaggaagcactgagcagctacagctaatgtgagtcgcagtcgtagcgcaaa 251
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1256 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1197

QY 252 gctatgggtagctacatcatggtgcatagctagcttcgtctgtagcagcatgtgcca 311
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1196 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1137
```

```
QY 312 tacctgtacacaattgccaacaacatcgatactgacgcatcaccacaattccaaca 371
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1136 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1077

QY 372 aacatacgaaccacaaa 388
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 1076 RRRRRRRRRRATCGCAA 1060
```

```
RESULT 3
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Szostak, Richard W.
; APPLICANT: Liu, Rhee
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

```
Query Match      7.9%; Score 30.6; DB 4; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.22;
Matches 14; Conservative 112; Mismatches 124; Indels 1; Gaps 1;

QY 21 agagagtgctcgtgtgcatcatgagctggaacacctacagaataccagggcgagcca 80
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 12 atarurarcrcrurururururarcrcrarururarcrcrarurgrnrnrnrnr 71

QY 81 tgcgagcagcagaacagcaactcgtgtgtaacgcaagctgtcgcaagtagaagctacgc 140
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 72 nrsnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 130

QY 141 gggcctggaagcagtgctacccagcagcagctgcccgcgagtaggaagcag 200
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 131 nrsnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 190

QY 201 gactgagcgtagctcagcagctaatgtgagtcgcagtcgctcgtagcagaagctatggg 260
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 191 nrsnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 250

QY 261 gtacactcata 271
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
DB 251 rgrcgrcrua 261
```

```
RESULT 4
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,932
FILING DATE: 16-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI42
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 932 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
;
US-08-242-932-7

Query Match          7.6%; Score 29.6; DB 1; Length 932;
Best Local Similarity 64.7%; Pred. No. 0.88;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

Db      605  ACAAATAATGATGAACAATATGATTCATGATCATTTATTATTAAGATTAGAAATTAATTTTAC 664
OY      380  gaaccaa 387
          ||| | ||
Db      665  GAAACTAA 672

RESULT 6
: Sequence 7, Application US/08714481
: Patent No. 5766606
: GENERAL INFORMATION:
: APPLICANT: Brady, L. Jeannine
: TITLE OF INVENTION: Cloning of No. 5766606-iga Fc Binding Forms of
: TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/714,481
: FILING DATE: 16-SEP-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION NUMBER:
: APPLICATION NUMBER: US 08/242,932
: FILING DATE: 16-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: UFI142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 7:
: LENGTH: 932 base pairs

```



```

; LOCATION: 307..311
; FEATURE:
; NAME/KEY: protein_bind
; LOCATION: 887..1507
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /bound_moiety="Iga Fc"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="Human Iga-Fc binding"
; OTHER INFORMATION: /label=Iga-binding
US-08-430-024-1

Query Match
Best Local Similarity 64.7%; Pred. No.1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgctgcaaacatgatactgacgcatcaccacaatttccaaaacaacataac 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ACAAAATTTGATGAACAATGATCTGATGATATTAGAAATTAATCAATTAAAC 787
OY 380 gaacacaa 387
    ||| ||| |||
Db 788 GAAACTAA 795

RESULT 9
US-08-782-009-1
; Sequence 1, Application US/08782009
; Patent No. 5714334
; GENERAL INFORMATION:
; APPLICANT: Faulmann, Ervin L
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
; TITLE OF INVENTION: AN IGA BINDING PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,009
; FILING DATE: 07-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,024
; FILING DATE: 27-APR-1995
; APPLICATION NUMBER: US 07/677,209
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: BL-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus agalactiae
```

```

; STRAIN: DL471
; IMMEDIATE SOURCE:
; CLONE: pELF26
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 320..430
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 320..1510
; OTHER INFORMATION: /codon_start=320
; OTHER INFORMATION: /function="binds to Fc region of human Iga"
; OTHER INFORMATION: /product="Iga binding protein"
; OTHER INFORMATION: /number=1
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 307..311
; FEATURE:
; NAME/KEY: protein_bind
; LOCATION: 887..1507
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /bound_moiety="Iga Fc"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="Human Iga-Fc binding"
; OTHER INFORMATION: /label=Iga-binding
US-08-782-009-1

Query Match
Best Local Similarity 64.7%; Pred. No.1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgctgcaaacatgatactgacgcatcaccacaatttccaaaacaacataac 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ACAAAATTTGATGAACAATGATCTGATGATATTAGAAATTAATCAATTAAAC 787
OY 380 gaacacaa 387
    ||| ||| |||
Db 788 GAAACTAA 795

RESULT 10
US-09-017-302-1
; Sequence 1, Application US/09017302
; Patent No. 6075128
; GENERAL INFORMATION:
; APPLICANT: Faulmann, Ervin L
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
; TITLE OF INVENTION: AN IGA BINDING PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/017,302
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/782,009
; FILING DATE: 07-JAN-1997
; APPLICATION NUMBER: US 08/430,024
; FILING DATE: 27-APR-1995
; APPLICATION NUMBER: US 07/677,209
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
```

```

NAME: Saliwanchik, David R
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: BL-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2617 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus agalactiae
STRAIN: DL471
IMMEDIATE SOURCE:
CLONE: pELF26
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 320..430
FEATURE:
NAME/KEY: CDS
LOCATION: 320..1510
OTHER INFORMATION: /codon_start=320
OTHER INFORMATION: /function="binds to Fc region of human IgA"
OTHER INFORMATION: /product="IgA binding protein"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: RBS
LOCATION: 307..311
FEATURE:
NAME/KEY: protein_bind
LOCATION: 887..1507
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /bound_moiety="IgA Fc"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="human IgA-Fc binding"
OTHER INFORMATION: /label="IgA binding"
/OS-09-017-302-1

```

```

Query Match      7.6%: Score 29.6; DB 3, Length 2617;
Best Local Similarity 64.7%: Pred. No.1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0.

QY      320  acacaatgtcgcaacacatcgatcactgacgcatccacaattccaaacaaacataac 379
      ||| ||| ||||| ||| ||||| ||||| ||| ||| ||| ||| |||
Db       728  ACAAAATTTGATGAAACAAATGATTCGTGATCATTTAGAAATTAAGAAATCAATTTAAC 787

QY      380  gaaccaaa 387
      ||| |||
Db       788  GAAACTAA 795

RESULT      11
US-08-923-992A-7
: Sequence 7, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-1gA Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA

```

```

1      ZIP: 20005
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: PatentIn Release #1.0, Version
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/08/923,992A
9      FILING DATE: 05-SEP-1997
10     CLASSIFICATION: 536
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 60/024,707
13     FILING DATE: 06-SEP-1996
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Esmond, Robert W.
16     REGISTRATION NUMBER: 32,893
17     REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (202) 371-2600
20     TELEFAX: (202) 371-2540
21     INFORMATION FOR SEQ. ID NO. 7:
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 3294 base pairs
24     TYPE: nucleic acid
25     STRANDEDNESS: double
26     TOPOLOGY: linear
27     MOLECULE TYPE: cDNA
28     FEATURE:
29     NAME/KEY: CDS
30     LOCATION: 1..3294
31     US-08-923-992A-7

```

US-08-923-992A-7

	Query Match	Best Local Similarity	Score 29.6%	DB 4	Length 3294
Matches	44	Conservative	0	Mismatches	24
				Indels	0
				Gaps	0
OY	320	acacatgctgcgaacacatcgatctacatccagatcaccacaaattccacaacacataac	379		
Db	316	ACAAAATTGATGAACACAAATGATTCGTGTCATTTTGAATTAGAAATCAATTATAC	375		
OY	380	gaaccaaa	387		
Db	376	GAAACTAA	383		

RESULT 12
 US-08-923-992A-3
 Sequence 3, Application US/08923992A
 Patent No. 6280738
 GENERAL INFORMATION:
 APPLICANT: Tai, Joseph Y.
 APPLICANT: Blake, Milan S.
 TITLE OF INVENTION: No. 6280738-1gA FC Binding Forms of the Group B
 TITLE OF INVENTION: Streptococcal Beta Antigens
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, N.W., Suite 600
 City: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:


```

APPLICANT: Nishikura, Kazuo
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,459
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/280,443
FILING DATE: 25-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Maly E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WSI49CUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6671 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 155..3832
US-08-457-459-1
Query Match 11.3%; Score 71.4; DB 1; Length 6671.
Best Local Similarity 80.0%; Pred. No. 1.3e-06;

```


RESULT 15

Search completed: January 17, 2002, 05:58:21
Job time: 1029 sec


```

: GENERAL INFORMATION:
: APPLICANT: Comer, Timothy W.
: APPLICANT: Wu, Kunsheng
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(5195)B
: CURRENT APPLICATION NUMBER: US/09/865,419A
: CURRENT FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/208,063
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ ID NOS: 54020
: SEQ ID NO 3373
: LENGTH: 381
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3236-040-P1-N1-D4
: US-09-865-419A-3373

```

Query Match	15.1%;	Score 95;	DB 5;	Length 381;
Best Local Similarity	62.3%;	Pred. NO. 8.9e-06;		
Matches 149;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0;

[illegible]

```

RESULT      3
US-09-865-439A-91961
; Sequence 91961, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51936)B
; CURRENT APPLICATION NUMBER: US/09/865.439A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/207,458
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 91961
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3607-057-Q6-N6-H9
; US-09-865-439A-91961

```

Query Match	14.5%;	Score 91.8;	DB 5;	Length 567;
Best Local Similarity	61.5%;	Pred. No. 2.6e-05;		
Matches 147; Conservative	0;	Mismatches 92;	Indels 0;	Gaps 0;

[illegible]

Qy	67	ttttttttttttttttttttttaaacaccccccatttttttagagagagccccc	126
Db	81	ttttttttttttttttttttttttaacatttttttttttttataaaaaaaaaa	140
Qy	127	acaaaagggggggcgctgtaaaaaacaactatgcttgcccttgggggtcaaccac	186
Db	141	aaaatatgggggggggggggaaaaaaaaccccacaatttccccacataaaaa	200
Qy	187	gggagtagttttttctgtaaaaaaaaacagggagtgagataccacacaaaaga	245
Db	201	aaaaaaaaaaaaaaaaaaagaaaggtatataaaaaaaaaaaaaaaaaaaaaaa	259

RESULT 4
US-09-696-664A-14530
; Sequence 14530, Application US/096966664A

```

1  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
2  TITLE OF INVENTION: Plants
3  FILE REFERENCE: 38-21(51721)P
4  CURRENT APPLICATION NUMBER: US/09/656,664A
5  CURRENT FILING DATE: 2000-10-25
6  PRIOR APPLICATION NUMBER: US 60/161,619
7  PRIOR FILING DATE: 1999-10-26
8  NUMBER OF SEQ ID NOS: 17472
9  SEQ ID NO 14530
10 LENGTH: 515
11 TYPE: DNA
12 ORGANISM: Zea mays
13 FEATURE:
14 NAME/KEY: unsure
15 LOCATION: (1)..(515)
16 OTHER INFORMATION: unsure at all n locations
17 OTHER INFORMATION: Clone ID: LIB3279-217-06-N6-B10
18 US-09-656-664A-14530

```

Query Match	14.58;	Score 91.6;	DB 5;	length 515;
Best Local Similarity	60.48;	Pred. No. 2.8e-05;		
Matches 151; Conservative	0;	Mismatches 99;	Indels 0;	Gaps 0;

[illegible]

RESULT 5
US-09-865-419A-36424
: Sequence 36424, Application US/09865419A
: GENERAL INFORMATION:
: APPLICANT: Conner, Timothy W.
: APPLICANT: Wu, Kunsheng

Query Match	14.3%	Score 90;	DB 5;	Length 602;
Best Local Similarity	72.2%;	Pred. No.	4.9e-05;	
Matches 117; Conservative	0;	Mismatches	45;	Indels 0;
Gaps	0;			

```

RESULT      8
US-09-865-439A-97784
: Sequence 97784, Application US/09865439A
: GENERAL INFORMATION:
: APPLICANT: Edgerton, Michael D
: APPLICANT: Hardeman, Kristine J.
: APPLICANT: La Rosa, Thomas J.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(51936)B
: CURRENT APPLICATION NUMBER: US/09/865,439A
: CURRENT FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/207,458
: PRIOR FILING DATE: 2000-05-30
: NUMBER OF SEQ ID NOS: 119126
: SEQ ID NO 97784
:
: LENGTH: 566

```

```

; TYPE:DNA
; ORGANISM: zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3608-021-N6-H10
US-09-B65-4339A-97784

```

Query Match	14.1%	Score	89.2	DB	5	Length	566
Best Local Similarity	57.0%	Pred.	No.6.4e-05				
Matches	163	Conservative	0	Mismatches	123	Indels	0
						Gaps	0

Qy 127 acaaaaaaggggggcygtgtaaaaaacaactcatgtgyccttcgggggtcaaccca 186
 ||||| || ||||| || ||||| ||
 Db 150 ttctggggggggggggggaaaaaaaaaaaaaaaaaaaacggggggggaaga 209

OY 247 gaggggtggaagagaggaatcccaaaagatgcagagaaga 292
| | | | | | | | | | | | | | | |
Db 270 aaaaagaaaaaaaaaaaaaaaaaaaaaaccaactcaaaaaa 315

RESULT 9
US-09-865-419A-3556

```

; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy W.
; APPLICANT: Wu, Kunsheng

```

```

: TITLE OF INVENTION:  Platts
:
: FILE REFERENCE:  38-21(51935)B
:
: CURRENT APPLICATION NUMBER:  US/09/865,419A
:

```

```

; PRIOR APPLICATION NUMBER: US 60/208,063
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 54020

```

```

; LENGTH: 424
; TYPE: DNA
; ORGANISM: Zea mays

```

OTHER INFORMATION: Clome ID: LIB3236-042-P1-N1-D12
US-09-865-419A-3556

Query Match	14.1%	Score 89;	DB 5;	Length 424;
Best Local Similarity	52.3%	Pred. No. 6.8e-05;		
Matches 197; Conservative	0;	Mismatches 180;	Indels 0;	Gaps 0;

```

Qy 67 tttttttttttttttttttaaacaccccccttatatttttagcagcagcccc 126
      ||||||| |||||
Db 73 ttttttttttttttttttttttttttttttttttttttttttttttttaa 132

```

QY 187 ggaggtagtttttcyatatgaaaaaacagygagygatataccaccacaaaaagag 246

Db 133 ttaaaaaaaagggggggaaaaaaaaccccttttttttaaaaaaaat 252

QY 247 gaggggctggcagagagagaaatccccaagaatgcgagagaccccatltgtga 306

Db 253 tttttttaaaaaaaagggccccaataaaatcttaaaaaaatttggggg 312

Qy 307 tgaagcttccccacacacccaagcataatgtgtgtgctgtcacaacaacatccccctatt 366
| |||| | | | | | | | | | | | | | |
Db 313 gcccttccccccaacaaaaaaatttatgtgggccccttacaagcccaaat 372

Qy	367	aaaaagagggacatcag	383
Db	373	tttaaaaaacgtatcgc	389

RESULT 10
US-09-865-439A-91290

```

; GENERAL INFORMATION:
;
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
;

```

```

; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/207,458
; PRIOR FILING DATE: 2000-05-30

```

```

; SEQ ID NO 912900
; LENGTH: 502
; TYPE: DNA

```

; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3607-052-Q6-N6-B9
 US-09-865-439A-91250

Query Match	14.1%;	Score	89;	DB	5;	Length	502;
Best Local Similarity	64.0%;	Pred.	No.	6.8e-05;			

QY 127 acacaaaggggggcygtggaacaaacacatcatgtgaccttcggggtcaaccaa 186
||||||| | ||||| | | ||||| ||| | ||||| ||| ||
127 4cc 100

RESULT 11
TTC-00-00E-420A-112007

```

; Sequence 112967, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; INVENTOR: Edgerton, Michael D
; ATTORNEY: Edgerton, Michael D

```

APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and other Molecules Associated With
TITLE OF INVENTION: Plants
TITLE OF INVENTION: Plants

; CURRENT APPLICATION NUMBER: US/09/865,439A
 ; CURRENT FILING DATE: 2001-05-29
 ; PRIOR APPLICATION NUMBER: US 60/207,458

[illegible]

		14.1%;	Score 88.8;	DB 5;	Length 503;
		Best Local Similarity 62.7%;	Po 7.3e-05;		
	Matches	138;	Conservative	0;	Mismatches 82; Indels 0; Gaps 0;
Oy	7				
Db	22				
Oy	67				
Db	82				
Oy	127				
Db	142				
Oy	187				
Db	202				

```

RESULT 14
US-09-865-419A-500
: Sequence 500. Application US/09865419A
: GENERAL INFORMATION:
: APPLICANT: Comer, Timothy W.
: APPLICANT: Wu, Kunsheng
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(51935)B
: CURRENT APPLICATION NUMBER: US/09/865,419A
: CURRENT FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/208,063
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ ID NOS: 54020
: SEQ ID NO 500
: LENGTH: 372
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3236-006-P1-N1-E6

```


[illegible][illegible]

RESULT	4
LOCUS	A1636456
DEFINITION	t6160456 293 bp mRNA EST 26-APR-1999
ACCESSION	AF1636456
VERSION	1
KEYWORDS	sequence.
SOURCE	similar to contains TARI.b2 MERR22 repetitive element ; , mRNA
ORGANISM	A1636456.1 GI:4687786
REFERENCE	EST.
AUTHORS	human.
TITLE	Homo sapiens
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
	1 (bases 1 to 293)
	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.
	Email: cgaps-remail.nih.gov
	Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
	Emmert-Buck, M.D., Ph.D.
	CDNA Library Preparation: Life Technologies, Inc.
	CDNA Library Arrayed by: Greg Lennon, Ph.D.
	DNA Sequencing by: Washington University Genome Sequencing Center
	clone distribution: NCI-CCAP clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at:
	www-bio.llnl.gov/brp/image/image.html
	Seq primer: -40UP from GIDCO
	High quality sequence stop: 283.
FEATURES	Location/Qualifiers
SOURCE	1..293

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2233059"
/clone_1b="NCI CGAP k168"
/tissue_type="renal cell tumor"
/lab_host="DH108"
/mote="Organ: kidney; Vector: pCMV-Sport6; Site_1: SalI"

```

Site_2: NotI; Cloned unidirectionally. Primer: 011g0 dT
Average insert size 1.2 kb. Life Technologies catalog #:
11524-014"

```

RESULT      5
A1493248
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
    1 (bases 1 to 735)
AUTHORS
    NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
JOURNAL
    Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-remail.nih.gov
COMMENT
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmert-Buck, M.D., Ph.D.
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CCGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LIML at:
    www-bio.liml.gov/bhrp/image/image.html
    Insert length: 1468 Std Error: 0.00
    Seq primer: ~400p from C10c0
    High quality sequence stop: 304.
    Location/Qualifiers
        1..735
FEATURES
    source

```

```

"organism":"Homo sapiens"
"db_xref":"taxon:9606"
"clone": "IMAGE:2131991"
"clone_1b": "NCI-CGAP_Gas4"
"tissue_type": "Poorly differentiated adenocarcinoma with
signal ring cell features"
"lat_host": "DH10B"
"note": "Organ: stomach; Vector: pCMV-SPORT6; Site: 1; Salt:

```


[illegible]

FEATURES	Location/Qualifi�ers
source	1. .425

BASE COUNT	ORIGIN
89 a	98 c 85 g 150 t 3 others

RESULT	11			
LOCUS	A1630928			
DEFINITION	A1630928	308 bp	mRNA	EST 26-APR-1999
ACCESSION	U99902.x1		NCI-CGAP U1 Homo sapiens cDNA clone IMAGE:2287226	3'
VERSION	A1630928		Similar to contains element MSRI Repetitive element ;	mRNA
KEYWORDS	sequence.			
	A1630928.1	GI:4682258		
	EST.			

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 308)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapsb@remail.nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLNt at: www-bio.linn.gov/bdipr/image/image.html Seq primer: -40UP from G1bco High quality sequence stop: 306. Location/Qualifiers 1..308
FEATURES	
SOURCE	

```

FEATURES
SOURCE
Location/Qualifiers
1..308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2287226"
/clone_idb="NCI_CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10b"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: Nott; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Life Technologies catalog #:
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:46:48 ; Search time 218.99 Seconds
(without alignments)
422.810 Million cell updates/sec

Title: US-09-823-101-6
Perfect score: 108
Sequence: 1 accatttctatatgtgaga.....aaaaaaaaaaaagcgcg 108

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_1101:*

1:	/SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.6	70.9	247	21	AAC04513
2	75	69.4	1133	20	AAZ24435
3	75	69.4	5145	21	AAC98196
4	61.8	57.2	982	22	AAZ21878
5	61.8	57.2	38682	22	AAZ21770
6	59.4	55.0	555	22	AAH33403
7	54.2	50.2	3265	21	AAZ65095
8	54.2	50.2	3265	22	AAZ62092
9	54.2	50.2	3265	22	AAZ64241
10	52	48.1	359	22	AAH83313
11	51.4	47.6	476	22	AAH83221

12	51.2	47.4	2802	20	AAZ56243
13	51.2	47.4	2910	20	AAZ56242
14	51	47.2	1052	21	AAZ63358
15	50.8	47.0	1064	21	AAZ36918
16	50.2	46.5	1580	22	AAI63862
17	50.2	46.5	1931	20	AAZ00477
18	50.2	46.5	1932	20	AAZ00411
19	50.2	46.5	1951	22	AAI63833
20	49.8	46.1	251	22	AAI57596
21	49.2	45.6	665	22	AAH34083
22	49.2	45.6	1153	20	AAV55748
23	49	45.4	1354	21	AAZ79921
24	49	45.4	3527	21	AAC99092
25	49	45.4	3648	21	AAC75860
26	48.8	45.2	887	21	AAC78089
27	48.8	45.2	2226	21	AAC77975
28	48.8	45.2	4189	21	AAC77054
29	48.6	45.0	1355	22	AAH34544
30	48.6	45.0	1640	20	AAI57594
31	48.6	45.0	2630	22	AAH33733
32	48.4	44.8	797	21	AAC79717
33	48.4	44.8	1080	21	AAC69612
34	48.4	44.8	3138	12	AAQ11712
35	48.2	44.6	772	22	AAZ32766
36	48.2	44.6	785	21	AAC77913
37	48.2	44.6	1181	19	AAV59803
38	48.2	44.6	1212	19	AAV59686
39	48.2	44.6	3060	22	AAZ02924
40	48.2	44.6	3060	22	AAZ44263
41	48.2	44.6	3060	22	AAZ44263
42	48.2	44.6	4257	20	AAZ98030
43	48	44.4	404	22	AAZ64787
44	48	44.4	894	20	AAV97723
45	48	44.4	997	20	AAZ52274

ALIGNMENTS

RESULT 1

AAZ04513	standard; cDNA: 247 BP.
ID	AAC04513;
AC	AAC04513;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein 5' EST, SEQ ID NO: 8588.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PF	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST) GENSET.
XX	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
DR	WPI: 2000-500381/45.
XX	
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX	diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS	Claim 1; SEQ ID 8588; 71pp + CD-ROM; English.

Human vitamin D re
Human vitamin D re
Human secreted pro
DNA encoding xenop
Human polynucleoti
Human secreted pro
Human secreted pro
Human polynucleoti
Human colorectal c
Human colon cancer
Human secreted pro
Human cancer assoc
Human pancreatic c
Human ORF1415
Human cancer assoc
Human cancer assoc
Human ORF2609
Human colon cancer
Human prostate can
Human colon cancer
Human secreted pro
Shuttle vector PMU
Human secreted pro
Human cancer assoc
Human secreted pro
Human PRO5723 cDNA
Human PRO5723 nucl
Human PRO5723 cDNA
Nucleic acid encod
Novel human polyu
Human adult retina
Protein PRO244 CDN

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or poly(A)⁺ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX
SQ Sequence 247 BP; 88 A; 37 C; 33 G; 88 T; 1 other;

Query Match 70.9%; Score 76.6; DB 21; Length 247;
Best Local Similarity 97.8%; Pred. No. 0.00031;
Matches 87; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataaa-tgtgaaaaaatgaataaagcatgttggc 64
|||||
Db 159 ttgtatattgagatgttcaataaatgtgaaaaaatgaataaagcatgttggc 218
|||||

QY 65 ttccaaagaagaaaaaa 93
|||||

Db 219 ttccaaagaagaaaaaa 247
|||||

RESULT 2
AAZ24435
ID AAZ24435 standard; cDNA; 1133 BP.
XX
AC AAZ24435;
XX
DT 14-FEB-2000 (first entry)
XX
DE Human bladder tumour cDNA library derived EST 47.
XX
KW Expressed sequence tag; human; bladder; tumour; cancer; cytostatic;
KW treatment; gene therapy; EST; ss.
XX
OS Homo sapiens.
XX
PN DE19818619-A1.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1998; 98DE-1018619.
XX
PR 21-APR-1998; 98DE-1018619.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-612028/53.
XX
PT New nucleic acid sequences expressed in bladder tumor tissue, and
PT derived polypeptides, for treatment of bladder tumor and identification
PT of therapeutic agents -
XX
PS Claim 3; Page 96; 132pp; German.

CC This invention describes novel polypeptide fragments (I) and the
CC polynucleotides (II) that encode them that are highly expressed in a
CC human bladder tumor and which have cytostatic activity. (II) are used
CC for recombinant expression of (I) and to isolate complete genes. (I) are
CC used to identify agents suitable for treatment of bladder cancer, to
CC directly treat this form of cancer (including expression from gene

CC therapy vectors) or are used in a preparation for cancer treatment. (I)
CC is also used for the generation of specific antibodies. (II) are
CC identified by assembling ESTs (expressed sequence tags) from a
CC particular tissue type before comparison of expression patterns. This
CC allows a significantly longer fragment of the gene to be revealed, and
CC therefore reduces the number of failures associated with the fact that
CC ESTs from different libraries may represent different parts of the same
CC unknown gene, distorting the estimated frequency of occurrence in a
CC particular tissue. AA43260-243309 represent expressed sequence tag (EST)
CC fragments isolated from a human bladder tumour cDNA library which encode
CC the proteins represented in AA46143-Y66138.

XX
SQ Sequence 1133 BP; 383 A; 198 C; 179 G; 373 T; 0 other;

Query Match 69.4%; Score 75; DB 20; Length 1133;
Best Local Similarity 98.9%; Pred. No. 0.00041;
Matches 86; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataaa-tgtgaaaaaatgaataaagcatgttggc 64
|||||
Db 1040 ttgtatattgagatgttcaataaatgtgaaaaaatgaataaagcatgttggc 1099
|||||

QY 65 ttccaaagaagaaaaaa 91
|||||

Db 1100 ttccaaagaagaaaaaa 1126
|||||

RESULT 3
AAC98196
ID AAC98196 standard; cDNA; 5145 BP.
XX
AC AAC98196;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:206.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardiovascular; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
XX
DR WPI; 2000-587534/55.
XX
PT P-PADB; AAB53439.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 1; Page 629-631; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardiovascular, muscular,
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnerable, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AA054007 represent sequences used in the exemplification of the present
XX invention.

SQ Sequence 5145 BP; 1099 A; 1302 C; 1499 G; 1236 T; 9 other;

Query Match 69.4%; Score 75; DB 21; Length 5145;
Best Local Similarity 98.9%; Pred. No. 0.00031;
Matches 86; Conservative 0; Mismatches 0; Indels 1; Gaps

DY 6 ttgtatattgtagagttttaataaa-tgtgaaaaaaatgaataagaacatttggtt 64
|||||
Db 5006 ttgtatattgtagagttttaataaatctgtgaataaacatgaataagaactttggtt 5065
|||||
DY 65 ttccaaaaagaaaaaaaaaaaaa 91
|||||
Db 5066 ttccaaaaagaaaaaaaaaaaaa 5092
|||||

RESULT 4
ID AAS21878 standard; DNA; 982 BP.
XX AAS21878:
AC AAS21878:
XX 24-OCt-2001 (first entry)
DT 24-OCt-2001 (first entry)
DE Human collagen gene COL1A2 3' UTR.
XX
XX Human: collagen: COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
KW osteoporosis: multiple epiphyseal dysplasia; osteogenesis imperfecta;
KM shortness of stature; low bone density; gene therapy; 3' UTR;
untranslated region.
XX
XX Homo sapiens.
OS
XX US6265157-B1.
PN 24-JUL-2001.
PD 24-JUL-2001.
PF 03-OCt-1997; 97US-0943731.
XX
PR 03-DEC-1991; 91US-0803628.
PR 13-MAR-1994; 94US-0212322.
XX
XX (UYAL-) UNIV ALLEGHENY HEALTH SCI.
PA (UYJ-) UNIV JEFFERSON THOMAS.
PA (UYOU-) UNIV ODU.
XX
XX Prockop DJ, Spottila LD, Deltas CD, Sereida L, Westerhausen Larson A;
PI Pack M, Collige A, Early J, Koerkoe J, Ala-Rokko L, Annunen S;
PI Pihlajamaa T, Vuoristo M, Paassilta P;
XX
DR WPI: 2001-432201/46.
XX
XX
XX Example 4; Fig 5B; 617pp; English.

The invention relates to Detecting a collagen gene alteration associated
with a pathological condition in a human subject by obtaining from the

CC	subject a sample nucleic acid containing a portion of at least 15
CC	consecutive nucleotides of the segment of the COL1A1 gene extending in
CC	the 5' to 3' direction from 78 nucleotides of Intron 27 located adjacent
CC	exon 28 through the 3' end of Intron 51, where the portion contains an
CC	intronic nucleotide and a first and second site, determining the sequence
CC	of the portion and comparing the sequence of the portion with the
CC	corresponding consensus sequence of the COL1A1 gene where a difference
CC	between the sequence of the portion and the consensus sequence indicates
CC	the presence of the collagen alteration in the subject. The method is
CC	used for detecting abnormalities in a COLI or COLII gene is useful for
CC	determining whether a subject is afflicted with pathological conditions
CC	associated with an altered collagen gene such as osteoporosis, multiple
CC	epiphyseal dysplasia, osteogenesis imperfecta, shortness of stature and
CC	low bone density. Identification of an abnormality in a collagen gene is
CC	also useful for designing a therapeutic nucleotide or gene therapy agent
CC	which can be administered to the subject to correct or alleviate the
CC	abnormality. The method is useful for detecting mutations in both the
CC	coding and non-coding sequences of any of the COLI or COLII genes.
CC	Therefore the method can be used to detect collagen gene alterations
CC	which affect either the primary sequence of a collagen protein chain,
CC	splicing of the mRNA encoding such chains or regulation of expression of
CC	the genes encoding such chains. The present sequence is the 3' UTR
CC	(untranslated region) of a collagen gene of the invention.
XX	
SQ	Sequence 982 BP; 323 A; 169 C; 159 G; 331 T; 0 other;
	Query Match 57.2%; Score 61.8; DB 22; Length 982;
	Best Local Similarity 90.6%; Pred. No. 0.041;
	Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps
OY	6 ttctgatatgtgagatgttctaataaa-tgtgaaaaaaatgaataagaatgttggtt 64
Dd	771 ttctgatattgtagagtgttttaataaatgtgaaaaaaatgaataagaatgttggtt 830
OY	65 ttccaaaagaaaaaaaaaaaaa 89
Dd	831 ttccaaaagacacatcgttagtaaaa 855
RESULT 5	
ID	AAS21770
AC	AAS21770 standard; DNA; 38682 BP.
XX	
AC	AAS21770;
XX	
DT	24-OCT-2001 (first entry)
DE	
XX	Human gene for collagen COL1A2.
XX	
KW	Human; collagen: COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
KW	osteoporosis; multiple epiphyseal dysplasia; osteogenesis imperfecta;
KW	shortness of stature; low bone density; gene therapy.
XX	
OS	Homo sapiens.
PN	US6265157-B1.
PD	
XX	24-JUL-2001.
PF	
XX	03-OCT-1997; 970S-0943731.
PR	
XX	03-DEC-1991; 91US-0803628.
PR	13-MAR-1994; 94US-0212322.
XX	
PA	(UYAL-) UNIV ALLEGHENY HEALTH SCI.
PA	(UYJE-) UNIV JEFFERSON THOMAS.
PA	(UYJU-) UNIV OULU.
PX	
PX	Prockop DJ, Spoltis LD, Deltas CD, Sereda L, Westerhausen Larson A;
PX	Pack M, Collige A, Early J, Koerkoe J, Ala-kokko L, Annunen S;
PX	Pinlajamaa T, Vuoristo M, Paasilta P;

DR WPI: 2001-432201/46.

XX Detecting collagen gene alteration, useful for diagnosing osteoporosis,
PT multiple epiphyseal dysplasia, osteogenesis imperfecta, shortness of
PR stature and low bone density in humans -
XX
PS Example 4; Fig 5A; 617pp; English.

XX The invention relates to detecting a collagen gene alteration associated
CC with a pathological condition in a human subject by obtaining from the
CC subject a sample nucleic acid containing a portion of at least 15
CC consecutive nucleotides of the segment of the COL1A1 gene extending in
CC the 5' to 3' direction from 78 nucleotides of Intron 27 located adjacent
CC exon 28 through the 3' end of Intron 51, where the portion contains an
CC intronic nucleotide and a first and second site, determining the sequence
CC of the portion and comparing the sequence of the portion with the
CC corresponding consensus sequence of the COL1A1 gene where a difference
CC between the sequence of the portion and the consensus sequence indicates
CC the presence of the collagen alteration in the subject. The method is
CC used for detecting abnormalities in a COL1 or COL9 gene is useful for
CC determining whether a subject is afflicted with pathological conditions
CC associated with an altered collagen gene such as osteoporosis, multiple
CC epiphyseal dysplasia, osteogenesis imperfecta, shortness of stature and
CC low bone density. Identification of an abnormality in a collagen gene is
CC also useful for designing a therapeutic nucleotide or gene therapy agent
CC which can be administered to the subject to correct or alleviate the
CC abnormality. The method is useful for detecting mutations in both the
CC coding and non-coding sequences of any of the COL1 or COL9 genes.
CC Therefore the method can be used to detect collagen gene alterations
CC which affect either the primary sequence of a collagen protein chain,
CC splicing of the mRNA encoding such chains or regulation of expression of
CC the genes encoding such chains. The present sequence is a collagen gene
CC of the invention.

XX
SQ Sequence 36882 BP; 11718 A; 7500 C; 7305 G; 12159 T; 0 other;

Query Match 57.2%; Score 61.8; DB 22; Length 36882;
Best Local Similarity 90.6%; Pred. No. 0.021;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaaataaa-tgtgaaaaaatagaataagcatgtttggtt 64
|||||
DB 38471 ttgtatattgagatgtttaaataaa-tgtgaaaaaatagaataagcatgtttggtt 38530
|||||

QY 65 ttccaaaagaataaaataaa 89
|||||
DB 38531 ttccaaaagaataatattgagtaaaa 38555
|||||

RESULT 6
AAH33403
ID AAH33403 standard; cDNA; 555 BP.
XX
AC AAH33403;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:459.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 7; ss.
XX
OS Homo sapiens.
XX
PN MO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI: 2001-235357/24.
DR
DR P-PSDB; AAG73972.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 2555; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 555 BP; 168 A; 92 C; 93 G; 193 T; 9 other;

Query Match 55.0%; Score 59.4; DB 22; Length 555;
Best Local Similarity 97.3%; Pred. No. 0.11;
Matches 71; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaaataaa-tgtgaaaaaatagaataagcatgtttggtt 64
|||||
DB 483 ttgtatattgagatgtttaaataaa-tgtgaaaaaatagaataagcatgtttggtt 542
|||||

QY 65 ttccaaaagaataa 77
|||||
DB 543 ttccaaaagaataa 555
|||||

RESULT 7
AAZ65095
ID AAZ65095 standard; cDNA; 3265 BP.
XX
AC AAZ65095;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO1124 encoding cDNA.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX
OS Homo sapiens.
XX
PN WO9963088-A2.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99MO-US12252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.

PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.

PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 20-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

XX
PA (GETH) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR P-PSDB; AY66749.
XX
XX
XX Membrane-bound proteins and related nucleotide sequences -
PS Claim 2; Fig 273; 822pp; English.
XX

CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.

SQ Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;

Query Match 50.2%; Score 54.2; DB 21; Length 3265;
Best Local Similarity 71.7%; Pred. No. 0.47;

Matches 71; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 5 attgtatattgtagatgtttaataatgtgaaaaaataagaataagcatgtttgtc 64

DB 3129 aatcattctatctgtagtgcataataacagtaagagagcaataacacatttgaa 3188

QY 65 ttccaaagaagaaaaaataaaaaaataaaaaa 103

DB 3189 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3227

RESULT 8

AAAF92092 ID AAF92092 standard; cDNA; 3265 BP.

AC AAF92092;

DT 15-MAY-2001 (first entry)

DE Human PRO1124 cDNA.

KW Human; PRO protein; mapping; ss.

OS Homo sapiens.

PN WO200116318-A2.

XX 08-MAR-2001.

PF 24-AUG-2000; 2000WO-US23328.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 07-DEC-1999; 99US-0169495.

PR 09-DEC-1999; 99US-0170262.

PR 11-JAN-2000; 2000US-0175481.

PR 18-FEB-2000; 2000WO-US04341.

PR 22-FEB-2000; 2000WO-US04414.

PR 01-MAR-2000; 2000WO-US05601.

PR 03-MAR-2000; 2000US-0187202.

PR 25-APR-2000; 2000US-0199397.

PR 22-MAY-2000; 2000WO-US14042.

PR 05-JUN-2000; 2000US-0209832.

XX (GETH) GENENTECH INC.

PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
DR MPI: 2001-183260/18.
DR P-PSDB; AAB87560.
XX

PT Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.

PS Claim 2; Fig 69; 278pp; English.

XX The present sequence is the coding sequence for a human PRO polypeptide
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC antagonists or anti-PRO antibodies are useful for preparation of a
CC medication useful in the treatment of a condition which is responsive to
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC protein may also be employed as molecular weight markers for protein
CC electrophoresis. The PRO coding sequence has applications in molecular
CC biology, including use as hybridisation probes, and in chromosome and
CC gene mapping.

SQ Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;

Query Match 50.2%; Score 54.2; DB 22; Length 3265;
Best Local Similarity 71.7%; Pred. No. 0.47;

Matches 71; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 5 attgtatattgtagatgtttaataatgtgaaaaaataagaataagcatgtttgtc 64

DB 3129 aatcattctatctgtagtgcataataacagtaagagagcaataacacatttgaa 3188

QY 65 ttccaaagaagaaaaaataaaaaaataaaaaa 103

DB 3189 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3227

RESULT 9

AAAF44241 ID AAF44241 standard; cDNA; 3265 BP.

AC AAF44241;

DT 02-APR-2001 (first entry)

DE Human PRO1124 (UNQ562) nucleotide sequence SEQ ID NO:378.

KW Human; secreted and transmembrane protein; PRO; cytosolic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay; ss.

OS Homo sapiens.

PN WO200073454-A1.

XX 07-DEC-2000.

PF 30-MAR-2000; 2000WO-US08439.

PR 02-JUN-1999; 99WO-US12252.

PR 23-JUN-1999; 99US-0141037.

PR 07-JUL-1999; 99US-0143048.

PR 20-JUL-1999; 99US-0144748.

PR 26-JUL-1999; 99US-0145698.

PR 28-JUL-1999; 99US-0146322.

PR 17-AUG-1999; 99US-0149396.

PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.

PR 08-OCT-1999; 99US-0158663.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28301.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 22-FEB-2000; 2000WO-US04414.


```

PR    24-FEB-2000; 200OWO-US04914.
PR    24-FEB-2000; 200OWO-US05004.
PR    02-MAR-2000; 200OWO-US05841.
PR    15-MAR-2000; 200OWO-US06884.
PR    20-MAR-2000; 200OWO-US07377.
XX
PA      (GETH ) GENENTECH INC.
XX
PI    Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL;
PI    Ferrara N, Fong S, Gerber H, Gertschen ME, Goddard A, Godowski PJ;
PI    Grimaldi CJ, Guirey AL, Kljavin IJ, Napier MA, Pan J, Poni NF;
PI    Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI    Zhang Z;
XX
DR    WPI; 2001-032160/04.
DR    P-PADB; AAB65272.
XX
PT    PRO polynucleotides used to produce polypeptides used to target
PT    bioactive molecules such as toxins, radiolabels or antibodies, to
PT    specific cells, to cause targeted cell death -
XX
PS    Claim 2; Fig 273; 935pp; English.
XX
CC    The present invention describes human secreted and transmembrane PRO
CC    proteins. The PRO proteins have cytosolic activity. The PRO proteins
CC    can be used for targeted delivery of bioactive molecules, such as
CC    toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC    sequences, and their fragments, can be used as hybridisation probes, in
CC    chromosomal and gene mapping, and in the generation of anti-sense RNA
CC    and DNA. They may also be used to produce transgenic animals which are
CC    used to develop and screen therapeutically useful reagents. The PRO
CC    nucleotide and protein sequence can be used for tissue typing and in
CC    treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC    AAFF4270 to AAFA4470 represent PCR primers and hybridization probes used
CC    in the isolation of human PRO sequences. AAFA4087 to AAFA4263 and
CC    AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC    sequences given in the exemplification of the present invention.
XX
SQ    Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other:
XX
Query Match          50.2%; Score 54.2; DB 22; Length 3265;
Best local Similarity 71.7%; Pred. No. 0.47;
Matches   71; Conservative     0; Mismatches   28; Indels     0; Gaps
QY       5 attgtatagtgcagtgcttaataaatgtaaaaaaaagaataaacgatgttggctt 64
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3129 aaatcatcattctggtgtagcaaatataaggtaaaggagacaataacacactttggaa 3188
QY       65 ttccaagaagaaaaaaaaaaaaaaaaaaaaaaaaaaa 103
        ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 3189 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3227
XX
RESULT_10
ID      AAH83313/c
AC      AAH83313 standard; cDNA; 359 BP.
XX
DT      25-SEP-2001 (first entry)
XX
DE      Human ovarian tumour associated polynucleotide sequence SEQ ID NO:937.
XX
KW      Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
KW      immunogenic; vaccine; ss.
XX
OS      Homo sapiens.
XX
PN      WO200151513-A2.
XX
DZ      19-JUL-2001.

```

PF	16-JAN-2001; 2001WO-US01575.	
XX		
PR	14-JAN-2000; 2000US-0176722.	
XX		
PA	(CORI-) CORIXA CORP.	
XX		
PI	Algate PA;	
XX		
DR	WPI; 2001-425866/45.	
PT	Novel ovarian tumor proteins, and nucleic acids encoding them, used to	
PT	treat and diagnose cancers, particularly ovarian cancer -	
XX		
PS	Claim 5; Page 235; 338pp; English.	
XX		
CC	AAH82377 to AAH83878 represent human ovarian tumor-associated	
CC	polynucleotide sequences which encode ovarian tumor proteins. The	
CC	ovarian tumor protein and polynucleotide sequences have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. The	
CC	ovarian tumor proteins and polynucleotides can be used to inhibit	
CC	the development of cancer, particularly ovarian cancer. They can also	
CC	be used to diagnose the onset and progression of cancer.	
XX		
SO	Sequence 359 BP; 120 A; 69 C; 73 G; 97 T; 0 other;	
XX		
Query Match	48.1%; Score 52; DB 22; Length 359;	
Best Local Similarity	98.4%; Pred. No. 1.5;	
Matches	63; Conservative 0; Mismatches 0; Indels 1; Gaps	
QY	6 ttgtatcatgltgagatglttaataaa-tgtgaaaaaaatgaataaagcatgttggc 64	
Db	64 TTGTATATGTGAGATGCTTTAAATAATTGTGAAAAAATGAATAAAGCATGTTGCTT 5	
QY	65 ttcc 68	
Db	4 TTCC 1	
XX		
RESULT 11		
AAH83221/c		
ID	AAH83221 standard; cDNA; 476 BP.	
XX		
AC	AAH83221;	
XX		
DT	25-SEP-2001 (first entry)	
XX		
DE	Human ovarian tumour associated polynucleotide sequence SEQ ID NO:845.	
KW	Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;	
KW	immunogenic; vaccine; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200151513-A2.	
XX		
PD	19-JUL-2001.	
XX		
XX	16-JAN-2001; 2001WO-US01575.	
PF		
XX	14-JAN-2000; 2000US-0176722.	
PR		
XX	(CORI-) CORIXA CORP.	
PA		
XX	Algate PA;	
PI		
XX	WPI; 2001-425866/45.	
DR		
PT	Novel ovarian tumor proteins, and nucleic acids encoding them, used to	
PT	treat and diagnose cancers, particularly ovarian cancer -	
XX		
PS	Claim 5; Page 219; 338pp; English.	
XX		

CC AAH8377 to AAH8378 represent human ovarian tumour-associated
CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.

XX
SQ Sequence 476 BP; 142 A; 97 C; 110 G; 126 T; 1 other;

Query Match 47.6%; Score 51.4; DB 22; Length 476;
Best Local Similarity 96.9%; Pred. No. 1.8;
Matches 63; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgtatcgtgagatcttaataaa-tgtgaataaatgaataagcatgttggtt 64
DB 178 TTTGTATATGTGAGATGTTTAAATTAATTGTGAAAAAATGAATTAAGCATGTTTGGTT 119
QY 65 ttcca 69
DB 118 TTCCA 114

RESULT 12

AA56243
ID AAX56243 standard; CDNA; 2802 BP.

XX
AC AAX56243;

DT 16-JUL-1999 (first entry)

XX Human vitamin D receptor related gamma 2 protein encoding cDNA.

DE Human: vitamin D receptor related protein; VDR; obesity; diabetes;
KW anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
KW hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
KW hyperproliferative skin disorder; hyperthyroidism; ss.

OS Homo sapiens.

XX
PN WO919354-A1.

PD 22-APR-1999.

PF 31-AUG-1998; 98WO-SE01548.

XX
PR 31-MAR-1998; 98SE-0001148.

PR 14-OCT-1997; 97SE-0003745.

XX (PHAA) PHARMACIA & UPJOHN AB.

PI Berkenstam A, Dahlberg M;

XX WPI: 1999-302508/25.

DR P-PSDB; AAY09516.

XX New vitamin D receptor related (VDR) polypeptides, useful for
PT treating obesity, diabetes, anorexia and rheumatoid arthritis

XX
PS Claim 2: Page 21-22; 35pp; English.

XX The present sequence encodes a human vitamin D receptor related (VDR)
CC polypeptide. Human VDR polypeptides and substances which affect VDR
CC signal transduction, can be used for treating metabolic, proliferative
CC or inflammatory conditions. They can be used in the manufacture of a
CC medicament for treating the following conditions: obesity, diabetes,
CC anorexia, lipoprotein defects, hyperlipidaemia, hypercholesterolaemia or
CC hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and
CC malign tumours, hyperproliferative skin disorders or hyperthyroidism.
CC Nucleic acid vectors encoding for expression of a VDR polypeptide can
CC be used for treating metabolic, proliferative or inflammatory conditions,
CC by introducing them into a mammal. The introduced nucleic acid is then

CC capable of transforming a cell in vivo and then polypeptide is expressed.
CC A substance affecting VDR signal transduction, such as an agonist or
CC antagonist can be used for the manufacture of a medicament for treating
CC metabolic, proliferative or inflammatory condition.
CC N.B. The specification specifically claims the VDR nucleic acid and
CC polypeptide sequences given in figures 1, 4, 7 and 8, but no figures
CC are given in the specification.

XX
SQ Sequence 2802 BP; 723 A; 715 C; 755 G; 609 T; 0 other;

Query Match 47.4%; Score 51.2; DB 20; Length 2802;
Best Local Similarity 77.5%; Pred. No. 1.4;
Matches 62; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 24 ttaataatctgtaaaaaaatgaataagcatgttggttccaaaagaaaaaaa 83
DB 2714 tcaaatcaaggcaaaaagaaatgaataatgactcttggcctaaaaa 2773
QY 84 aaaaaaaaaaaaaaaaaa 103
DB 2774 aaaaaaaaaaaaaaaaaa 2793

RESULT 13

AA56242
ID AAX56242 standard; CDNA; 2910 BP.

XX
AC AAX56242;

DT 16-JUL-1999 (first entry)

XX Human vitamin D receptor related gamma protein encoding cDNA.

DE Human: vitamin D receptor related protein; VDR; obesity; diabetes;
KW anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
KW hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
KW hyperproliferative skin disorder; hyperthyroidism; ss.

OS Homo sapiens.

XX
PN WO919354-A1.

PD 22-APR-1999.

PF 31-AUG-1998; 98WO-SE01548.

XX
PR 31-MAR-1998; 98SE-0001148.

PR 14-OCT-1997; 97SE-0003745.

XX (PHAA) PHARMACIA & UPJOHN AB.

PI Berkenstam A, Dahlberg M;

XX WPI: 1999-302508/25.

DR P-PSDB; AAY09515.

XX New vitamin D receptor related (VDR) polypeptides, useful for
PT treating obesity, diabetes, anorexia and rheumatoid arthritis

XX
PS Claim 2: Page 17-18; 35pp; English.

XX The present sequence encodes a human vitamin D receptor related (VDR)
CC polypeptide. Human VDR polypeptides and substances which affect VDR
CC signal transduction, can be used for treating metabolic, proliferative
CC or inflammatory conditions. They can be used in the manufacture of a
CC medicament for treating the following conditions: obesity, diabetes,
CC anorexia, lipoprotein defects, hyperlipidaemia, hypercholesterolaemia or
CC hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and
CC malign tumours, hyperproliferative skin disorders or hyperthyroidism.
CC Nucleic acid vectors encoding for expression of a VDR polypeptide can
CC be used for treating metabolic, proliferative or inflammatory conditions,
CC by introducing them into a mammal. The introduced nucleic acid is then

CC	capable of transforming a cell in vivo and then polypeptide is expressed.
CC	A substance affecting VDR signal transduction, such as an agonist or
CC	antagonist can be used for the manufacture of a medicament for treating
CC	metabolic, proliferative or inflammatory condition.
CC	N.B. The specification specifically claims the VDR nucleic acid and
CC	polypeptide sequences given in figures 1, 4, 7 and 8, but no figures
CC	are given in the specification.
XX	
SQ	Sequence 2910 BP; 765 A; 728 C; 780 G; 636 T; 1 other;
XX	
Query Match	47.4%; Score 51.2; DB 20; Length 2910;
Best Local Similarity	77.5%; Pred. No. 1.4;
Matches	62; Conservative 0; Mismatches 18; Indels 0; Gaps 0
OY	24 ttaataaatgctgaataaagcatgtttgttcccaagaagaaaaa 83
DB	2818 tcaatcaaggcacaaggacttaataatgcttcttgctcaaaaaa 2877
OY	84 aaaaaaaaaaaaaaaaaa 103
DB	2878 aaaaaaaaaaaaaaaaaa 2897
RESULT 14	
AAA26358	
ID	AAA26358 standard; cDNA; 1052 BP.
AC	AAA26358;
XX	
DT	29-JUN-2000 (first entry)
XX	
DE	Human secreted protein gene 13 SEQ ID NO:23.
XX	
KW	Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW	antiHIV; antiInflammatory; nootropic; neuroprotective; antiAllergic;
KW	osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW	antiporiatic; cardiant; gene therapy; cancer; neurological disorder;
KW	immune disease; inflammation; blood disorder; tumour; chromosome 16; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200006698-A1.
PN	
PD	10-FEB-2000.
XX	
PF	29-JUL-1999; 99WO-US17130.
XX	
XX	30-JUL-1998; 98US-0094657.
PR	05-AUG-1998; 98US-0095486.
PR	06-AUG-1998; 98US-0095454.
PR	06-AUG-1998; 98US-0095455.
PR	12-AUG-1998; 98US-0096319.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI	Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI	Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX	
XX	WPI: 2000-195282/17.
DR	P-PSDB: AAY91463.
XX	
PT	New isolated human genes and the secreted polypeptides they encode,
PT	useful for diagnosis and treatment of e.g. cancers, neurological
PT	disorders, immune diseases, inflammation or blood disorders -
XX	
XX	Claim 1; Page 380-381; 634pp; English.
XX	
CC	The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC	human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC	proteins can have activities based on the tissues and cells they are
CC	expressed in. Examples of the activities are: cytostatic;

CC	Immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC	antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC	antislamic; antipsoriatic; and cardiant. The polynucleotides and their
CC	corresponding secreted proteins are useful for preventing, treating or
CC	ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC	pathological conditions can be diagnosed by determining the amount of the
CC	proteins in a sample or by determining the presence of mutations in the
CC	polynucleotides. Specific uses are described for each of the
CC	polynucleotides, based on which tissues they are most highly expressed
CC	in, and include developing products for the diagnosis or treatment of
CC	cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC	and foetal deficiencies, blood disorders, diseases of the immune system,
CC	autoimmune diseases, hepatic and renal disease, inflammation,
CC	allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC	osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC	transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC	cardiovascular disorders, reproductive disorders, gastrointestinal
CC	disorders, respiratory disorders and metabolic disorders. The proteins
CC	or polynucleotides can also be used as food additives or preservatives.
CC	The proteins are also useful for identifying their binding partners.
CC	AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC	exemplification of the present invention.
SO	
XX	Sequence 1052 BP; 265 A; 294 C; 312 G; 181 T; 0 other;
XX	
XX	Query Match 47.2%; Score 51; DB 21; Length 1052;
XX	Best Local Similarity 85.1%; Pred. No. 1.7; Mismatches 10; Indels 0; Gaps 0
XX	Matches 57; Conservative 0; Mismatches 10; Indels 0; Gaps 0
Qy	37 aaaaatcgaataaagcatgttgcgtttccaaaagaaaaaagaaaaaagaaaaa 96
Db	
	936 aaaaaataaaaaaataaaatttttaatttttaaaaaaataaaaaaataaaaaa 995
Qy	97 aaaaaa 103
Db	996 aaaaaa 1002
RESULT 15	
AAZ36918	standard: DNA; 1064 BP.
XX	
AC	AAZ36918;
XX	
DT	13-MAR-2000 (first entry)
XX	
DE	DNA encoding Xenopus Geminin-H protein.
XX	
DE	Nuclear protein; Geminin-L; DNA replication; neurogenesis; neural growth;
KW	Geminin-H; isoform; pre-replication complex; origin of replication;
KW	nuclear DNA replication; anaphase promoting complex; ectodermal cell;
KW	neural cell; neural hypertrophy; ectopic neurogenesis;
KW	future neural plate; early gastrula; bone morphogenetic protein 4; BMP4;
KW	proliferative disease; cancer; neurological disease; Parkinson's disease;
KW	Alzheimer's disease; multiple sclerosis; spinal cord injury; ss.
XX	
XX	Xenopus sp.
OS	
XX	
FH	Key Location/Qualifiers
FT	CDS 53..712
FT	/*tag= a
FT	/product= "Geminin-H"
XX	
PN	WO9958673-A1.
XX	
PD	18-NOV-1999.
XX	
PF	13-MAY-1999; 99WO-US10577.
XX	
PR	13-MAY-1998; 98US-0085371.
PR	11-JUN-1998; 98US-0096724.
XX	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 05:58:21 ; Search time 93.06 Seconds
(without alignments)
262.837 Million cell updates/sec

Title: US-09-823-101-6

Perfect score: 108
Sequence: 1 acctatttgatattgtgtaga.....aaaaaaaaaaagcgcg 108

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents: NA: *
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/lna/PCtUS.COMB.seq: *
6: /cgn2_6/ptodata/2/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61.8	57.2	982	4	US-08-943-731-110 Sequence 110, Appl
2	61.8	57.2	38682	4	US-08-943-731-2 Sequence 2, Appl
3	48.4	44.8	3138	1	US-07-867-106-4 Sequence 4, Appl
4	48	44.4	1604	1	US-08-665-966-9 Sequence 9, Appl
5	48	44.4	1604	3	US-09-041-780-9 Sequence 9, Appl
6	47.8	44.3	2255	4	US-08-871-572B-3 Sequence 3, Appl
7	47.2	43.7	1558	1	US-08-455-550-7 Sequence 7, Appl
8	47.2	43.7	5613	2	US-08-463-418-1 Sequence 1, Appl
9	46.8	43.3	3637	1	US-08-445-640-3 Sequence 3, Appl
10	46.8	43.3	3637	3	US-08-170-558-3 Sequence 3, Appl
11	46.8	43.3	3637	3	US-08-447-314-3 Sequence 3, Appl
12	46.8	43.3	3637	3	US-08-445-461-3 Sequence 3, Appl
13	46.8	43.3	4931	4	US-08-726-320-2 Sequence 2, Appl
14	46.8	43.3	4931	4	US-09-208-716-2 Sequence 2, Appl
15	46.8	43.3	5852	1	US-07-867-106-2 Sequence 2, Appl
16	46.6	43.1	688	6	5498694-3 Patent No. 5498694
17	46.4	43.0	127	3	US-09-014-416-59 Sequence 59, Appl
18	46.4	43.0	176	3	US-09-014-416-63 Sequence 63, Appl
19	46.4	43.0	183	3	US-09-014-416-60 Sequence 60, Appl
20	46.4	43.0	200	3	US-09-014-416-64 Sequence 64, Appl
21	46.4	43.0	9595	3	US-09-014-416-4 Sequence 4, Appl
22	46.4	43.0	9599	3	US-09-014-416-2 Sequence 2, Appl
23	46.4	43.0	9599	3	US-09-014-416-6 Sequence 6, Appl
24	46.2	42.8	742	1	US-07-847-010-12 Sequence 12, Appl
25	45.8	42.4	198	1	US-08-330-108-16 Sequence 16, Appl
26	45.8	42.4	198	5	PCR-US92-10087-16 Sequence 16, Appl
27	45.8	42.4	1332	2	US-09-057-762-1 Sequence 1, Appl

28	45.8	42.4	1332	3	US-08-326-119A-1 Sequence 1, Appl
29	45.8	42.4	3581	2	US-08-738-349-1 Sequence 1, Appl
30	45.6	42.2	1582	3	US-08-545-196B-10 Sequence 10, Appl
31	45.6	42.2	1582	3	US-08-545-196B-12 Sequence 12, Appl
32	45.4	42.0	1117	4	US-09-247-373B-33 Sequence 33, Appl
33	45.4	42.0	1559	2	US-08-417-174-1 Sequence 1, Appl
34	45.4	42.0	1559	2	US-08-231-565A-1 Sequence 1, Appl
35	45.4	42.0	1559	2	US-09-007-961-1 Sequence 1, Appl
36	45.4	42.0	1559	4	US-09-267-439-1 Sequence 1, Appl
37	45.2	41.9	1441	4	US-08-821-994-63 Sequence 63, Appl
38	45.2	41.9	1813	4	US-09-071-224-3 Sequence 3, Appl
39	45.2	41.9	3410	4	US-09-020-956-110 Sequence 110, Appl
40	45.2	41.9	3410	4	US-09-030-607-110 Sequence 110, Appl
41	45.2	41.9	3572	2	US-08-713-815A-2 Sequence 2, Appl
42	45	41.7	780	2	US-08-540-804-36 Sequence 36, Appl
43	45	41.7	780	4	US-08-590-399-36 Sequence 36, Appl
44	45	41.7	1639	4	US-09-362-473-5 Sequence 5, Appl
45	45	41.7	3848	4	US-09-112-096-28 Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-943-731-110
Sequence 110, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISSA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KOKKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PANTICH SCHWARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:
LENGTH: 982 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-110

Query Match 57.2%; Score 61.8; DB 4; Length 982;
Best Local Similarity 90.6%; Pred. No. 0.00037;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgtttggt 64
|||||
Db 771 TTTGTATATGTGAGATGTTTAAATAATTGTGAATAAATGAATGAATGAATGTTGTT 830

QY 65 ttccaaagaaaaaataaaaaaa 89
|||||
Db 831 TTCCAAAGAACATTTGAGTAA 855

RESULT 2

US-08-943-731-2

Sequence 2, Application US/08943731

Patent No. 6265157

GENERAL INFORMATION:

APPLICANT: PROCKOP, DARWIN J.

APPLICANT: SPOTILA, LORETTA D.

APPLICANT: DELTAS, CONSTANTINOS D.

APPLICANT: SEREDA, LARISA

APPLICANT: LARSON, ANDREA W.

APPLICANT: PACK, MICHAEL

APPLICANT: COLIGE, ALAIN

APPLICANT: EARLY, JAMES

APPLICANT: KORRKO, JARMO

APPLICANT: ALA-KORRKO, LERNA, et al.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING

NUMBER OF SEQUENCES: 666

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.

STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND

STREET: FLR.

CITY: PHILADELPHIA

STATE: PA

COUNTRY: USA

ZIP: 19103-7086

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,731

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,322

FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/803,628

FILING DATE: 03-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: DOYLE LEARY Ph.D., KATHRYN

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9598-27

TELEPHONE: 215-965-1284

TELEFAX: 215-567-2991

TELEX: 831-494

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 38682 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-2

Query Match 57.2%; Score 61.8; DB 4; Length 38682;
Best Local Similarity 90.6%; Pred. No. 0.00027;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgtttggt 64
|||||
Db 38471 TTTGTATATGTGAGATGTTTAAATAATTGTGAATAAATGAATGAATGTTGTT 38530

QY 65 ttccaaagaaaaaataaaaaaa 89
|||||
Db 38531 TTCCAAAGAACATTTGAGTAA 38555

RESULT 3

US-07-867-106-4/C

Sequence 4, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: SLADE, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved plasmid Vectors for Cellular

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rfs

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3439

TELEFAX: 215-568-3100

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3138 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

US-07-867-106-4

Query Match 44.8%; Score 48.4; DB 1; Length 3138;
Best Local Similarity 68.4%; Pred. No. 0.084;
Matches 67; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Oy	6	tttgatcagtcgaacgctttaaataaacgtgaaaaaaatcgaaataagaactgttgctt	65
Dd	2054	TTATATTAAATTATNTCTGTTTTAAGAAATAGAAAAAAAAAAAAAAAATCTTTTATGC	1995
Oy	66	tcccaaatgaataaaaaaaaaaaaaaaaaaaaaaaa	103
Dd	1994	AATCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1957

```

RESULT 4
US-08-665-966-9
: Sequence 9, Application US/08665966
: Patent No. 5756328
: GENERAL INFORMATION:
: APPLICANT: Steffens, John C.
: APPLICANT: Changas, Gurdev S.
: TITLE OF INVENTION: Acyl Transferase and Gene Encoding Acyl
: TITLE OF INVENTION: Transferase
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Jones, Tullar & Cooper, P.C.
: STREET: P.O. Box 2266 Eads Station
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/665,966
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spector, Eric S.
: REGISTRATION NUMBER: 22495
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-415-1500
: TELEFAX: 703-415-1508
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1604 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 35..1392
: US-08-665-966-9

Query Match 44.4%; Score 48; DB 1; Length 1604;
Best Local Similarity 68.8%; Pred. No. 0.11;
Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps

QY 8 tgtatcgtagatggttaataatcgtcaaaataaagaacatggttggttc 67
||||| ||||| || ||| ||||| || ||| ||||| || ||| |||||
Db 1490 TGTATCTTTCACGTAAATTCACACATTTAAAGATATATCATTCGAGTTGCGCTTA 1549

QY 68 caaaagaaaaaataaaaaaataaaaaaataaaaaa 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1585

RESULT 5
US-09-041-780-9
: Sequence 9, Application US/09041780
: Patent No. 6066482
: GENERAL INFORMATION:
: APPLICANT: Steffens, John C.

```

```

1  APPLICANT:  Changas, Gurdev S.
2  TITLE OF INVENTION:  Acyl Transferase and Gene Encoding Acyl
3  TITLE OF INVENTION:  Transferase
4  NUMBER OF SEQUENCES:  19
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  Jones, Tullar & Cooper, P.C.
7  STREET:  P.O. Box 2266 Eads Station
8  CITY:  Arlington
9  STATE:  Virginia
10 COUNTRY:  USA
11 ZIP:  22202
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  Patent In Release #1.0, Version #1.25
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER:  US/09/041,780
21
22 FILING DATE:
23
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:  08/665,966
27
28 FILING DATE:
29
30 ATTORNEY/AGENT INFORMATION:
31 NAME:  Spector, Eric S.
32 REGISTRATION NUMBER:  22495
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE:  703-415-1500
35 TELEFAX:  703-415-1508
36
37 INFORMATION FOR SEQ ID NO:  9:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH:  1604 base pairs
40 TYPE:  nucleic acid
41 STRANDEDNESS:  double
42 TOPOLOGY:  linear
43
44 MOLECULE TYPE:  CDNA
45
46 FEATURE:
47 NAME/KEY:  CDS
48 LOCATION:  55..1392
49
50 US-09-041-780-9

```

```

Query Match 44.4%; Score 48; DB 3; Length 1604;
Best Local Similarity 68.8%; Pred. No. 0.11;
Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 8 tgtatattgtagatgtttaataaataatgtaaaataatgaataaagcatgtttgttc 67
      ||||| ||||| ||| | ||||| ||||| ||||| |||||
Db 1490 tctatcttggacatgtataatcaccacatttataataatattatcattcgagttgacctta 1549

Qy 68 caaaagaataaataaataaataaataaataaataa 103
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1550 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1585

RESULT 6
US-08-871-572B-3
Sequence 3, Application US/08871572B
Patent No. 6287853
GENERAL INFORMATION:
APPLICANT: Pestka, Sidney
APPLICANT: Kotenko, Serguei
APPLICANT: Soh, Jaemog
APPLICANT: Donnelly, Robert
APPLICANT: Mariano, Thomas
APPLICANT: Cook, Jeffrey
APPLICANT: Emmanuel, Stuart
APPLICANT: Schwartz, Barbara
TITLE OF INVENTION: Accessory Factor for Interferon Gamma
TITLE OF INVENTION: and Its Receptor
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESS: Richard R. Muccino

```

```

      APPLICATION NUMBER: 08/056,564
      FILING DATE: 30-APR-1993
      APPLICATION NUMBER: 07/577,892
      FILING DATE: 05-SEP-1990
      ATTORNEY/AGENT INFORMATION:
      NAME: Eisenstein, Ronald I
      REGISTRATION NUMBER: 30628
      REFERENCE/DOCKET NUMBER: 40302-FWC-DIV
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-523-3400
      TELEFAX: 617-523-6440
      TELEX: 200291
      INFORMATION FOR SEQ ID NO: 7:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1558 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: CDNA
      HYPOHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE:
      ORIGINAL SOURCE:
      US-08-4550-7

Query Match      43.7%   Score 47.2; DB 1; Length 1558;
Best Local Similarity 72.6%; Pred. No. 0.15;
Matches 61; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
20 atgttaataaagctgtaaaataaagcatttggtttccaaagaaaaa 79

```

```

Query Match          43.7%; Score 47.2; DB 1; Length 1558;
Best Local Similarity 72.6%; Pred. No. 0.15;
Matches      61: Conservative    0; Mismatches   23; Indels     0; Gaps     0;

Oy       20 atgttaataaatcgtgaaaaaatgaaataagcatgttggltttccaaaagaanaa 79
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1450 ACGTATTAAAAAGAAATAAAAAAAAAATTAATTAAGTCATTATTTTAACAATAAAAAAA 1509
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy       80 aaaaaaaaaaaaaaaaaaaaaa 103
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db       1510 AAAAAAAAAAAAAAAAAAAAAA 1533
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      8
US-08-463-418-1/c
: Sequence 1, Application US/08463418
: Patent No. 5908971
GENERAL INFORMATION:
APPLICANT: Van Der Straeten, Dominique et al.
TITLE OF INVENTION: CRUCIFER ACC SYNTHASE AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,418
FILING DATE: 05-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/962,481
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/161002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
```



```

;
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-463-418-1

Query Match          43.7%; Score 47.2; DB 2; Length 5613;
Best Local Similarity 72.1%; Pred. No. 0.13;
Matches 75; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY 5 attgtatcgtgagatglttaataatgtaaaaaatgaataagcatgttggtt 64
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 2148 ATTATATTAGGCGTAAAGTTTATA-AAATAAGTACTAATTAATGCAACGAACTTT 2090

QY 65 ttccaaaagaaaaaagaaaaaagaaaaaagcgcg 108
    | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 2089 AACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAACACG 2046

RESULT 9
US-08-445-640-3
; Sequence 3, Application US/08445640
; Patent No. 5709858
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baron, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,640
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C2
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3637 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-445-640-3
```

```

Query Match          43.3%; Score 46.8; DB 1; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16;
Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 ttgtagatgtttaataatgtaaaaaatgaataagcatgttggttccaaaag 73
    ||| | | |||| | | | | | | | | | | | | | | | | | | | | |
Db 3539 TGAGGCAATTTTAAATCCCTGACTAGCGACGTAATAATAAGTTGAGTTTCACAAA 3598

QY 74 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3599 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3628

RESULT 10
US-08-170-558-3
; Sequence 3, Application US/08170558
; Patent No. 6001621
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baron, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/170,558
; FILING DATE: 20-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C1
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3637 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-170-558-3

Query Match          43.3%; Score 46.8; DB 3; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16;
Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 ttgtagatgtttaataatgtaaaaaatgaataagcatgttggttccaaaag 73
    ||| | | |||| | | | | | | | | | | | | | | | | | | | | |
Db 3539 TGAGGCAATTTTAAATCCCTGACTAGCGACGTAATAATAAGTTGAGTTTCACAAA 3598

QY 74 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726.320
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: AF-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-4166
TELEFAX: 415-845-0555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-726-320-2

Query Match
Best Local Similarity 43.3%; Score 46.8; DB 4; Length 4931;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 22 gtctaatgaatgtaaaatgaatgaatgctgtgtgtttccaaagaaaaaa 81
Db 4842 GGTTCCTACCTGTAATACCTGTAATGAATGATTGATTATTAATAAAAAA 4901
Qy 82 aaaaaaaaaaaaaaaaaa 103
Db 4902 AAAAAAAAAAAAAAAAAA 4923

RESULT 14
US-09-208-716-2
Sequence 2, Application US/09208716
Patent No. 6235880
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto.
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/208,716
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/726.320
FILING DATE:
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: AF-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-09-208-716-2

Query Match
Best Local Similarity 43.3%; Score 46.8; DB 4; Length 4931;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 22 gtctaatgaatgtaaaatgaatgaatgctgtgtgtttccaaagaaaaaa 81
Db 4842 GGTTCCTACCTGTAATACCTGTAATGAATGATTGATTATTAATAAAAAA 4901
Qy 82 aaaaaaaaaaaaaaaaaa 103
Db 4902 AAAAAAAAAAAAAAAAAA 4923

RESULT 15
US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5389526/r1s
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2
```

```
Query Match 43.3%; Score 46.8; DB 1; Length 5852;
Best Local Similarity 67.3%; Pred. No. 0.15;
Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
QY 6 ttgtatgtgagatgttaataatgtaaaaaatgaataaagcatgttgtt 65
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5697 TTTTATTATTAAATTAAATTAATTAAATTAAATTAATAAAAAA 5638

QY 66 tccaaagaaaaaataaaaaaataaaaaa 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5637 ATTTAAATTAAAAAATAAAAAAATAAAAAA 5600
```

Search completed: January 17, 2002, 05:58:24
Job time: 1032 sec

RESULT	1			
AX093215				
LOCUS				
DEFINITION	AX093215	396 bp	DNA	PAT
ACCESSION	Sequence 33	from Patent	WO0118046.	
VERSION	AX093215			
KEYWORDS	AX093215.1	GI:13509664		
SOURCE				
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
JOURNAL	1 (bases 1 to 396).			
	Xu,J. and Stolk,J.A.			
	Ovarian tumor sequences and methods of use therefor			
	Patent: WO 0118046-A 33 15-MAR-2001;			
	CORIMX CORPORATION (US)			
FEATURES	Location/Qualifiers			
Source	1..396			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			

[illegible]

FEATURES	SOURCE
COMMENT	Martinisried GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinisried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZ434C0820) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 10559 Berlin-Charlottenburg, GERMANY; Email: cloneetzp@d.further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/. Location/Qualifiers
gene	1..750 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZp434C0820" /clone_id="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="adult" /issue_type="testis" 1..444 /gene="DKFZp434C0820" <1..444 /note="Similarity to cyclin B3 - chicken" /codon_start=1 /product="hypothetical protein" /protein_id="CA870806.1" /db_xref="GI:6808235" /translation="LVNLKCDINPIAYHFPLRYARCILHTNMKTITLSRYICEWTLQEE YHYVQEKASKIAASALITLALYMKLGWGVPLEHSXGYSISELPYRLNKLTPSS YDSLKAVYYKSHPEVFVEVAKIRPDLMLKEIIINCCEAGGLV" polya_signal polysite BASE COUNT 265 a 147 c 126 g 212 t ORIGIN 651 .633
Query Match	Best Local Similarity 12.2%; Score 76.8; DB 9; Length 750; Matches 90; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
OY	7 ttggggaacctttt 66
Dd	736 ttt 677
OY	67 tttttttttttttttttttttaaacaccccccttatatttttagcaga 118
Dd	676 ttttttttttttttttttttttttttaacaatraccaatttaatttaaA 625
RESULT	6
LOCUS	AB063070.c
DEFINITION	Macaca fascicularis brain cDNA clone:QJRA-10484, full insert sequence.
ACCESSION	AB063070
VERSION	AB063070.1 GI:14388527
KEYWORDS	oligo capping; fts (full insert sequence).
SOURCE	Macaca fascicularis adult male temporal lobe right cDNA to mRNA, clone_1lib:macaque brain cDNA library QJRA clone:QJRA-10484. Macaca fascicularis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
ORGANISM	1 (sites)
REFERENCE	Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K. Isolation of full-length cDNA clones from macaque brain cDNA libraries
AUTHORS	Unpublished
TITLE	2 (bases 1 to 1555) Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S. Direct Submission

[illegible]

```

-----
Project Information
Center Project Name: 409883
Center clone name: RPCI-11_9C14
-----

Summary Statistics
Consensus quality: 64021 bases at least Q40
Consensus quality: 69210 bases at least Q30
Consensus quality: 69869 bases at least Q20
Estimated insert size: 150370; agarose-fp estimation
Estimated insert size: 73896; sum-of-contigs estimation
Quality coverage: 7.64 in Q20 bases; agarose-fp estimation
Quality coverage: 15.54 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a "working draft" sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

    1      1386: contig of 1386 bp in length
    *      1387      1486: gap of unknown length
    *      1487      2573: contig of 1087 bp in length
    *      2574      2673: gap of unknown length
    *      2674      4382: contig of 1709 bp in length
    *      4383      4482: gap of unknown length
    *      4483      10249: contig of 5767 bp in length
    *      10250     10349: gap of unknown length
    *      10350     15995: contig of 5646 bp in length
    *      15996     16095: gap of unknown length
    *      16096     19413: contig of 3318 bp in length
    *      19414     23239: contig of 3726 bp in length
    *      23240     23339: gap of unknown length
    *      23340     32199: contig of 8860 bp in length
    *      32200     32299: gap of unknown length
    *      32300     42648: contig of 10349 bp in length
    *      42649     42748: gap of unknown length
    *      42749     54536: contig of 11788 bp in length
    *      54537     54636: gap of unknown length
    *      54637     74896: contig of 20260 bp in length.

FEATURES
source          Location/Qualifiers
                1..74896
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone="RP11-9C14"
BASE COUNT      24523 a 13463 c 12999 g 22899 t 1012 others
ORIGIN
Query Match           12.2% Score 76.8; DB 2; Length 74896;
Best Local Similarity 70.8%; Pred. NO. 0.00021;
Matches 102; Conservat 0; Mismatches 42; Indels 0; Gaps 0;

Oy       7   ttctttttttttttttttttttttaatttctttggtggaaacctttt 66
            |||
Db        31288 tttyyyyyyyyyyyyyyyyyyytttttttttttttttttttttttttttttt 31229

Oy       67   ttctttttttttttttttttttaaacacccccccttatcttttagcagagcacccc 126
            |||
Db        31228 tttyyyyyyyyyyyyyyyyyyytttttttttttttttttttttttttttttttttcacacaccccc 31169

Oy       127   acaaaaagggggggcggtgaag 150
            |  |||
Db        31168 AACATTAACAAGCTTGAGGAA 31145

RESULT      8
HSZA242859/c 1999 bp mRNA PRI 02-DEC-1999
LOCUS

```


DEFINITION	Homo sapiens mRNA for langerin protein.
ACCESSION	AJ242859
VERSION	AJ242859.1 GI:5523278
KEYWORDS	langerin protein.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1999) Valladeau,J., Ravel,O., Dezutter-Dambuyant,C., Moore,K., Kleijmeer,M., Duvert-Frances,V., Vincent,C., Schmitt,D., Davoust,J., Caux,C., Lebecque,S. and Saeland,S. langerin, a new transmembrane C-type lectin specific to Langerhans cells, induces the formation of Birbeck granules Unpublished 2 (bases 1 to 1999) Valladeau,J.M. Direct Submission Submitted (02-JUN-1999) Valladeau J.M., Laboratory for Immunological Research, Schering-Plough, 27 Chemin des Peupliers, 69571 Dardilly Cedex, FRANCE Revised by author 06-SEP-1999 Location/Qualifiers 1..1999 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2" /cell_type="Langerhans cells" /map="2p13" 48..1034 /codon_start=-1 /product="langerin" /protein_id="CAB62403.1" /db_xref="GI:5523278" /translation="MTVEKKAPDAHPFYVDKONISLAMPREPPKSGPSLPVKETPTVR AIICTITVLVASYLQAVLVLPREMGTISDYKTNTQLAKGRDNISTDSELTPEKSDGK EAGVOIQOMYNESLGIVRSOFLKLKTSYERNAIOIILTRSWESTLNIAOIPETKRS LEKASALNTRIKRALOGSLNNMSKLKRQNDLLQVNSGMYKKGFFYSFSLPIKTVYS AEQFCVSRNSHLTSTSESDEQEFLYKTAGLIWIGLIKAMECDMSWDVTPEKVVS SARWIRGEPPNNAAGNEHCNIKAPSIQAMNDAPCDRTFLPICKRPVPSP"
CDS	polya_signal BASE COUNT 586 a 494 c 450 g 469 t ORIGIN
Query Match	12.1% Score 76.4; DB 9; Length 1999;
Best Local Similarity	59.0%; Pred. No. 0.00023;
Matches 131; Conservative	0; Mismatches 91; Indels 0; Gaps 0;
OY 7	tcttttctttctttctttctttctttctttaattctttcttgsggaaacctttt 66
Db 1973	TTTTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1914
OY 67	tcttttctttctttctttctttctttctaacaaccccccttatcttttagcagacgccccc 126
Db 1913	TTTTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1854
OY 127	acaaaaaggggggggcggtggaaaaacaactcatgtgccttcgggggtcaaacccoa 186
Db 1853	GCAATTTTAATTCTGGAGGAAGCTGGAAAGTGTAATTCATGGAAGCGCATGAAGCTGTCCCA 1794
OY 187	ggagtagcttctttgatgtaaaaaaacaacggaagaaga 228
Db 1793	TGCACAAGCACACAGAAGATGTAAGAAATTCACAGAAGAGAGGCA 1752
RESULT 9	
BC007326/c	BC007326 1195 bp mRNA PRI 12-JUL-2001
LOCUS	Homo saplens, clone MGC:14879 IMAGE:3617525, mRNA, complete cds.
DEFINITION	BC007326
ACCESSION	BC007326.1 GI:13938381
VERSION	MGC.
KEYWORDS	

[illegible]

```

SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 4670)
AUTHORS        Prasad,P.D., Wang,H., Huang,W., Kekuda,R., Rajan,D.P., Leibach,F.H.
               and Ganapathy,V.
TITLE          Human LAT1, a subunit of system L amino acid transporter: molecular
               cloning and transport function
JOURNAL        Biochem. Res. Commun. 255 (2), 283-288 (1999)
REFERENCE      2 (bases 1 to 4670)
AUTHORS        Prasad,P.D., Wang,H., Huang,W., Kekuda,R., Rajan,D.P., Leibach,F.H.
               and Ganapathy,V.
TITLE          Direct Submission
JOURNAL        Submitted (02-NOV-1998) Obstetrics and Gynecology, Medical College
               of Georgia, 1120 15th Street, Augusta, GA 30912, USA
FEATURES       Source
               Location/Qualifiers
                 1..4670
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /tissue_type="Placenta"
                   67..1590
                     /function="transports large neutral amino acids in
                     association with 4F2HC antigen"
                     /note="LAT1"
                     /codon_start=1
                     /product="L-type amino acid transporter subunit LAT1"
                     /protein_id="AAD20464.1"
                     /db_xref="GI:426640"
                   /translation="MACAGKRRRLAAPAEKEENAKMLAASADGAPAGEGEV
                   TLQNLITLNGVAIIIVGTIGSGIFVPYGLKAGSPGLAVYMAAGVFSTVGALC
                   VAEIGTITISKGGDYAYMLEVYGSLPAFLKIMLELIIRPSOYLVALVATYLLKP
                   PPTCPVEBAKKLVACLCVLILTVANCYSVAARVDGAFAALAILIILGFVO
                   IKGDSNLDLPNEFEGTKLDGVNIATLSXGAFRAGAMNVLTREEMINPYNPDL
                   AIIISLPVTLYVLYLTNAVPTTLSTDOMSSSEAAVDEGNYHGWMSWIIIPRVGIS
                   CFGVNSGLETSRSLPFGVSRGHLPIILSMIHOLIPLPDAIYPFCVMILYAFSD
                   IFESINFPENMCVALATIGMTWLNRHRDELRPRTKNVIALVFFITLACLFIAYS
                   FMKIPVEGCIFFTIISGLPFYFFGVMMKPKMWLOGIFSTVLCOKLMQVVPEET"
BASE COUNT     896 a 1471 c 1301 g 1002 t
ORIGIN
Query Match           12.1%; Score 76.2; DB 9; Length 4670;
Best Local Similarity 74.4%; Pred. No. 0.00025;
Matches    96; Conservative   0; Mismatches    33; Indels     0; Gaps     0;

QY      7 |tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt|
Db      4623 TTTT'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT' 4564

QY      67 |ttttttttttttttttttttttaaacacccttatatttttagcagcgacccc|
Db      4563 TTTT'TTTTTTTTTTTTTTTTTTTTAGATGATGTGAGTGCTTCCTTTAAAAATAAACCCCG| 4504

QY      127 acaaaaaaag 135
              ||||||||
Db      4503 ACAAAAAG 4495

RESULT  11
LOCUS    AX093192                396 bp    DNA
DEFINITION Sequence 10 from Patent WO0118046.
ACCESSION AX093192
VERSION   AX093192.1    GI:13509641
KEYWORDS
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS   Xu,J. and Stolk,J.A.
```

```

TITLE          Overlari tumor sequences and methods of use therefor
JOURNAL        Patent: WO 0118046-A 10-15-MAR-2001;
FEATURES       CORIXA CORPORATION (US)
SOURCE         Location/Qualifiers
               1..396 /organism="Homo sapiens"
misc_feature   1..396 /db_xref="taxon:9606"
BASE COUNT     85 a      18 c      55 g      153 t      85 others
ORIGIN
Query Match    12.0% Score 76; DB 6; Length 396;
Best Local Similarity 78.6%; Pred. No. 0.00026;
Matches 88; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Oy 7           |ttttttttttttttttttttttttt|taatttttttcgtgggaacacctttt 66
                |||||||||
Db 22          |TTTTTTT|TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 81
                |||||||||

Oy 67          |ttttttttttttttttttttaaacacccttatatccttagcacgaa 118
                |||||||||
Db 82          |TTTTTTTTTTTTTTTTTTTTTTTAATAAAAAANNNNTTTTTTTTTTTTMAAA 133
                |||||||||

RESULT 12
AK026528/c     AK026528 1483 bp mRNA PRI 29-SEP-2000
LOCUS          Homo sapIens CDNA: FLJ22875 fts, clone KAT02879.
DEFINITION     AK026528
ACCESSION      AK026528
VERSION        GI:104339405
KEYWORDS       oligo capling; fts (full insert sequence).
SOURCE         Homo sapIens signet-ring cell carcinoma cell_line:KATO III CDNA to
ORGANISM       mrna, clone.lib:KAT clone:KAT02879.
              Homo sapIens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
AUTHORS       Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
              Nakamura.Y., Isogai.T. and Sugano.S.
              NEBO human cDNA sequencing project
              Unpublished (2000)
              2 (bases 1 to 1483)
              Sugano,S., Suzuki,Y., Ota,T., Obayashi.M., Nishi,T., Isogai,T.,
              Shibahara,T., Tanaka,T. and Nakamura.Y.
              Direct Submission
              submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
              Sugano, Institute of Medical Science, University of Tokyo,
              Laboratory of Genome Structure Analysis, Human Genome Center;
              Shirokane-dai, 4-6-1, Minato-ku, TOKYO 108-8639, Japan
              (E-mail:cdaieims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
              Fax:81-3-5449-5416)
COMMENT        NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
constructing, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokoyo (partly supported by science and Technology
Agency)).
Location/Qualifiers
1..1483
/organism="Homo sapIens"
/db_xref="taxon:9606"
/cell_line="KATO_III"
/cell_type="signet-ring cell carcinoma"
/clone="KAT02879"
/clone_id="KAT"
/note_"cloning vector pWE18FLU3"
31..513
/note_"unnamed protein product"
/codon_start=1
```

[illegible]

ORGANISM	Homo sapiens
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE	MRNA, clone_11b:KAT clone:KATO6217.
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	1 (sites)
TITLE	Nakamura,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
JOURNAL	Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
AUTHORS	2 (bases 1 to 1163)
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
JOURNAL	Shibahara,T., Tanaka,T. and Nakamura,Y.
AUTHORS	Direct Submission
TITLE	Submitted (15-FEB-2000) to the DDBJ/EMBL/Genbank databases. Sunio
JOURNAL	Sugano, Institute of Medical Science, University of Tokyo, Department
AUTHORS	of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
TITLE	(E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
JOURNAL	Fax:81-3-5449-5416)
AUTHORS	NEO human cDNA sequencing project supported by Ministry of
TITLE	International Trade and Industry of Japan: cDNA full insert
JOURNAL	sequencing: Research Association for Biotechnology: cDNA library
AUTHORS	construction, 5'- and 3'-end one pass sequencing: Department of
TITLE	Virology and Human Genome Center, Institute of Medical Science,
JOURNAL	University of Tokyo (partly supported by Science and Technology
AUTHORS	Agency).
FEATURES	Location/Qualifiers
source	1..1163
CDS	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_line="KATO III"
	/cell_type="signet-ring cell carcinoma"
	/clone="KAT06217"
	/clone_id="KAT"
	/note="Cloning vector pME10SF13"
	49...699
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="BA91288.1"
	/db_xref="GI:7020826"
	/translation="MRRPSCPTRNPLRAFVGYTRGRTGQSRCAEAESACGERPA
	RFGAPAGTGWGAHSASEVELEGKTGFSSDQLEQLHRFKQLSGQPPIRKENFNN
	VPDLINIRSKIVRAFEIDNRLRKGPSGLADEINFDFLITMSYFPIIDTMDEGV
	ELSKREKRIFLFHYDSDSDSRITLLEEVNRNVKWSRCRETITSRSPLASPSPGP"
BASE COUNT	323 a 332 c 319 g 189 t
ORIGIN	
Query Match	12.0%; Score 75.6; DB 9; Length 1163;
Best Local Similarity	82.1%; Pred.No.0.00032;
Matches	87; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Qy	7 ttgggaaaaactttt 66
Ddb	1150 ttt 1091
Qy	67 tttttttttttttttttttttttaaacaccccccttatatttt 112
Ddb	1090 ttttttttttttttttttttttttaattaaacaaccttttttttttttt 1045
RESULT	15
LOCUS	HSMB00144/c
DEFINITION	Homo sapiens mRNA; CDNA DKFPZ566G013 (from clone DKFPZ566G013);
ACCESSION	AL050277
VERSION	AL050277.1 GI:4886454
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

THIS PAGE BLANK (USPTO)


```
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-182-816-26
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgatgcagcagcagatgtgtgtatgtctgtcttgcactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 480 ctgtgcaggtttgttagctgtgatgaaatctctgtatcttccaagaagggtgaaga 539
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 540 agtcttcgctcaa 552
```

```
RESULT 9
US-09-471-528-25/c
; Sequence 25, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-471-528-25
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgatgcagcagcagatgtgtgtatgtctgtcttgcactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 916 ctgtgtcaggtttgttagctgtgatgaaatctctgtatcttccaagaagggtgaaga 857
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 856 agtcttcgctcaa 844
```

```
RESULT 10
US-09-471-528-26
; Sequence 26, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-471-528-26
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgatgcagcagcagatgtgtgtatgtctgtcttgcactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 480 ctgtgcaggtttgttagctgtgatgaaatctctgtatcttccaagaagggtgaaga 539
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 540 agtcttcgctcaa 552
```

```
RESULT 11
US-09-634-530-25/c
; Sequence 25, Application US/09634530
; Patent No. 6290958
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/634,530
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 09/471,528
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 09/182,816
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 08/989,510
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-634-530-25
```

```
Query Match          23.8%; Score 25; DB 4; Length 1395;
```

```
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY 29 ctgtgtacgatgatgcagcagcagatgtgtgtatgtctgtcttgcactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
```


; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-471-528-22

Query Match 23.8%; Score 25; DB 3; Length 1736;
Best Local Similarity 58.9%; Pred. No. 5.3;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 29 ctgtgacgatagcagcagcagatgtgtatgtctgtctcttggaactaggatcagc 88
DB 1074 CTGTGTCAGCTTTGTAGCTTGATGGAATATCTCTGATCTTCCAAAGGCTGAGA 1015
QY 89 cgccttctctgaa 101
DB 1014 AGTGTTCGCTCAA 1002

Search completed: January 17, 2002, 05:58:26
Job time: 1034 sec


```
259 AAAAGCAGCAGGAAGAATATGCC.....AAATACTCA 293
      |||      ::      |||      :::::
129 aValAlaThrLysCysGlnAlaIaThrGluGluArgValAla 146
      :::::
294 CAGTGATTTGCTGAGTGGTGAATTATGGCTGATTCCTTGTCTTTTA 343
      ::|||  ||:::
146 IaValThrLeuAlaLys.....AlaGluAlaMetAlaPheLeu 158
      ::|||  ||:::
344 .....TATTCGAATTTT 357
      :::::
159 GlnGluGlnProPheLysAspPheValThrSerAlaPheTyrAspLysPh 175
      :::::
358 TATAAGA.....GCACAGAAATACTTTT 380
      :::::
175 eLeuGlnTrpLysLeuPheGluMetGlnProValSerAspLysTyrPheT 192
      ::|||  ||:::
381 AAGATTTCTGACTACAAAAATTGGTTTAAATCCCTGTGCCAAGTT 430
      |||||  ::|||  ::|||  ::|||  |||
192 hrgIuPheArgValLeuGlyLysGlyLysGlyGluValCysAlaVal 208
      ::|||  ::|||  ::|||  ::|||  |||
431 CAGAAATATGCCATAATGGAACAAGATACTAAAGAGAAGAACTACCTC 480
      |||  |||  |||  |||  |||
209 Gln.....ValLysAsnThrG1 214
      ::|||  |||  |||  |||  |||
481 AAGGTTA.....AAAAACCAAAAAAAGG 521
      ::|||  |||||  |||||  |||||  |||||
214 yLysMetTyrAlaCysLysLysLeuAspLysArgLeuLysLysG 231
      ::|||  |||||  |||||  |||||  |||||
522 GGGGGGA 529
      |||||  |||||  |||||  |||||  |||||
231 yGlyGlu 233
```


OM of: US-09-823-101-1 to: Issued_Patents_AA* out_format : pfs

Date: Jan 17, 2002 4:00 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-O/cgn2_1/USPTO.spool/US09823101/runat.17012002.155137_22014/app_query.fasta_1.2239
-DB=Issued_Patents_AA -QFMT=fasten -SUFFIX=n2p.ra1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=bloms62 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09823101 -ECGN1_1_59 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEXY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-1
Query length: 542
Database: Issued_Patents_AA*
Database sequences: 212252
Database length: 22503292
Search time (sec): 66.610000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation		
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-11				75.00	146.26	0.3063	351
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-19				73.50	133.97	0.5971	871
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-13				72.00	123.89	1.10	1724
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-221-817-13				70.50	131.21	1.29	576
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-454-439-13				70.50	131.21	1.29	576
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-947-947A-2				69.50	126.32	1.84	753
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-737-078A-2				68.50	116.17	1.22	47
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-464-954A-6				68.50	126.67	2.31	576
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-3				68.50	109.98	3.56	3174
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-947A-2				67.00	125.76	3.34	446
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-650-650-5				66.00	147.07	2.43	40
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-11				66.00	132.42	3.55	179
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-8				66.00	132.04	3.58	186
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-1				66.00	129.96	3.88	255
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-997-803-15				65.50	129.58	4.29	213
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-506-12				65.50	125.57	4.76	321
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-506-20				65.50	125.57	4.76	321
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-836-786-4				64.50	147.50	3.40	27
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-152-488-8				64.50	145.84	3.55	32
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-303-025-8				64.50	145.84	3.55	32
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-677-304-8				64.50	145.84	3.55	32
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-436-703B-13				64.50	145.84	3.55	32
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-183-959-12				64.50	125.21	6.06	264
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-6				64.50	100.92	11.39	3169
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-890A-58				64.00	141.38	4.47	45
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-58				64.00	141.38	4.47	45
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-971A-58				64.00	141.38	4.47	45
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-90A-60				64.00	138.73	4.79	59
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-60				64.00	138.73	4.79	59
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-971A-60				64.00	138.73	4.79	59
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-462-040-60				64.00	103.07	12.10	2265
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-149-097D-36				64.00	102.06	12.41	2509
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-8				64.00	99.69	13.21	3200
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-221-817-22				63.50	115.32	9.88	576
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-454-439-22				63.50	115.32	9.88	576
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-959-104-11				63.50	113.88	10.26	667

/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-7 - 63.50 98.29 15.38 3287
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-097-830E-2 + 63.00 143.07 5.40 30
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-456-112B-2 + 63.00 143.07 5.40 30
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-890A-62 + 63.00 131.30 7.33 100
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-62 + 63.00 131.30 7.33 100

seq_name: /cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-11

seq_documentation block:

Sequence 11, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 351
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: SITE
LOCATION: all Xaa positions
OTHER INFORMATION: Xaa-unknown amino acid
US-09-245-041-11

alignment_scores:
Quality: 75.00 Length: 29
Ratio: 3.261 Gaps: 0
Percent Similarity: 79.310 Percent Identity: 51.724

alignment_block:
US-09-823-101-1 x US-09-245-041-11 ..
Align seg 1/1 to: US-09-245-041-11 from: 1 to: 351

433 AAGATACTAAAGAGAACTACCTCAGGTTAAACCAACCAAAAA 502
|||||
24 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
503 AAAAAAAAAAAGGGGGGGAACAGGCA 539
|||||
40 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 52

seq_name: /cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-19

seq_documentation block:
Sequence 19, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 19
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-245-041-19

```

```

alignment_scores:
  Quality: 73.50      Length: 114
  Ratio: 1.246       Gaps: 9
  Percent Similarity: 51.754   Percent Identity: 29.825

```

```

alignment_block:
US-09-823-101-1 x US-09-245-041-19 ..

```

```

Align seg 1/1 to: US-09-245-041-19 from: 1 to: 871

```

```

194 TGTGTAGTGCACATCATGCTCAGTCCACATGAAAAATCCAGCTCCACT 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
792 CysValGlyProLeuLeuGlnProAlaSerProAsnThrValGlnProLy 808
244 TTTCATATATGCAGAAAAAGACAGAGAGAAATATGCCAAATATAGTC. 292
      :: :: :: :: ||| ||| ||| ||| ||| ||| ||| ||| |||
808 sleuAsnLeuAlaGlu.....GlyLysSerPheCysProPheIleP 822
293 .....ACAGTGAATTTGTCTGAGTGTGAATTATGCGCTGATTTCTGTC 337
      || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
822 roHISThrSerIleMetGly.....Phe 829
338 TTTTATATTTCCAAATTTTATATAGACACAGAAATACTTTTATAGAAAT 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
830 PheValAlaPheAsnAsnThrValLeu.....LysTyrLeu.....Ph 841
388 TCTGACTACA.....AAAATGTGTTTTTAAATTCCTGTCGACAG 428
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 eleuLeuSerPheGluIleLysAsnIleLeu.....CysCysSerV 855
429 TTCAGAAATATGCCATTAATGAAACAGATTAAGTAAAGAAAAACTACC 478
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
855 alLysLys.....LysLysLys.....LysLysLys..... 860
479 TCAAGGTTAAAAAACCAAAAAAAGAAAAAAG 520
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
861 .....LysLysLysLysLysLysLysLysLysLysLys 871

```

```

seq_name: /cgn2_6/ptodata/2/laa/5B_COMB.pep:US-08-477-451-15

```

```

seq_documentation_block:

```

```

; Sequence 15, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacgi, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113

```

```

; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1724 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-15

```

```

alignment_scores:
  Quality: 72.00      Length: 187
  Ratio: 0.986       Gaps: 9
  Percent Similarity: 39.037   Percent Identity: 22.460

```

```

alignment_block:
US-09-823-101-1/rev x US-08-477-451-15 ..

```

```

Align seg 1/1 to: US-08-477-451-15 from: 1 to: 1724

```

```

418 GGAATTAATAAACCAATTTT..... 398
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 GlyMetGlnLysProLeuPheArgAsnIleSerCysSerValLeuPheAl 1471
397 .....TGATGTCACA 388
1471 aleuPheCysArgSerSerAspArgTyrLeuArgLysTyrCysIleHis 1488
387 AATCTTAATAAGTATTTCTGTGCTGTATATAAATTTGAAATA..... 344
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1488 erPheLeuIleIlePheTrpIleGlyTyrGlnAsnSerCysLeuLysPro 1504
343 .....TA 342
1505 ThrAsnThrIlePheValPheValArgAsnSerIlePheAspTrpSerPh 1521
341 AAAAGAACCAATCAGCCAT..... 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1521 eSerGluTyrSerIleProHisHisLeuHisPheSerThrSerLysLysP 1538
319 ..AATTCACACATCAGACAAATACATGATGATATTTTGCATATTTTCCT 272
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1538 roHISerProLeuSerIleSerHisSerTyrGlnSerProCysThrPro 1554
271 TCC.....TGTCCTTTTCTGTGCATATAGAAAACTGG.. 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1555 SerValProSerLysAlaCysProPheGluGlnTyrLeuLeuLeuTrpse 1571
238 .AGAGCTGATTTTTCAGTGGCAGTGCATGAGTGAGCTACACAGAA 190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1571 rGlnThrCysPhePheGlnAlaValSerAsnAsnLeuTyrTyrProIleH 1588
189 TGTTGCTTTTTCATATTAGTGCATGTGCTCTGCGCTC..... 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1588 euPheCys.....SerThrAspLysCysTyrTrpAsn 1598
148 .....TGTCGCTTACCTTTCCSCCATAGTACCTCTCTTTTTC 108
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1599 IleLeuGlnSerLeuAlaCysLysIleLysLeuIleGlySerLeu.PheC 1615
107 TCTTTTTCSTA.....TTTGTCTTACCCCTTCC 79
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1615 yspPheValLeuHisLeuLysLysThrHisProPheGlnAlaLysProGlu 1631
78 CCTCAGCTTA 70
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1632 ProTyrLeu 1634

```

```

seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:US-08-221-817-13

```

```

seq_documentation_block:
; Sequence 13, Application US/08221817
; Patent No. 553151
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5532151el G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,817
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5532151and, Greta E.
; REGISTRATION NUMBER: 35,302
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEAX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-221-817-13

alignment_scores:
Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

alignment_block:
US-09-823-101-1 x US-08-221-817-13 ..
Align seg 1/1 to: US-08-221-817-13 from: 1 to: 576

42 ATGTGAACAAACAGACCACTCTTAAGTTAGTGAGGGGAAGGCTAGAA 91
:::|||||::: |||||::: ||||| |||
6 Tlevalaiaasntrhvalleuleulsalaarggluglglglglgla 22
::: ||| |||:::||||| |||:::
22 nargllyslly LysSerlysllystrparglmetleu..... 34
::: ||| ||||| ::| |||||
142 CCAGAGACGACGACAGACACATGACCTTAATATGAAAC...AGAAC 188
35 .....GlnPheProHisIleSerGlnCysglugluLeuArgle 47
169 ATTCTTGATGAGTCCACTCATGCTGCATGCCACATGAAAAATCCAGCTCT 238
||| |||::: |||::: |||::: ::| ::|:::
47 userleuGlunArgAspTyrHisSerLeuCysGlu...ArgHisArgIleG 63

```

```

239 CCAGTTTCTATATGACAG..... 258
::: |||::: |||
63 LysArgLeuLeuPheArgGluPheCysAlaThrArgProGluLeuSerArg 79
259 .....AAAAAGA 266
80 CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAspAs 96
267 CAGGAGGAAGAAATATGC.....CAAAAT.....AGTCACA 295
|:::|:::|::: ||| ||||| |||||
96 PLYsArgGlySAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296 GT..... 297
113 hrglyProAspLeuIleProGluValProArgGlnLeuValThrAsnGys 129
297 ..... 297
130 ThrGlnArgLeuGluGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
298 .....GATTTGTCTGAGTGTGATATGAGCTGATCTTG 334
146 rArgLeuThrHisGluTyrLeuSerValAlaProPheAlaAspTyrLeu 162
335 TTCCTTTATATTTCCAAATTTTATAAGACACAGAAATACTTTTAAGA 384
:::|:::|:::|::: |||:::|:::|:::
163 AspSerIleTyrPheAsnArgPheLeuGln...TrrpLysTrpLeuGluAr 178
385 ATTCTGACTACAAAAAATGTTT..... 409
178 glnProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
410 .....TTTAATTCCTGCGCAAGTTCAGAAATATGCCATATGSAACAA 454
195 LysGlyPheGlyGluValCysAlaCysGln..... 204
455 GATACTAAAGAAAGAACTACTCAAGTTA.....AAAAACC 495
205 .....ValArgAlaThrGlyLysMetTyrAlaCysLysLysLe 217
496 AAAAAAAAAAAAAAAAAAAGGGGGGGA 529
:::|:::|:::|::: |||:::|:::|:::
217 uGluLysLysArgIleLysLysArgLysGlyGlu 228

seq_name: /cgn2_6/prodata/2/iaa/5A_COMB.pep:US-08-454-439-13

seq_documentation_block:
; Sequence 13, Application US/08454439
; Patent No. 5591618
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5591618el G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,439
; FILING DATE: 30-MAY-1995

```

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,817
FILING DATE: 31-MAR-1994
APPLICATION NUMBER: 08/123,932
FILING DATE: 17 SEP 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 559161band, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31981
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-454-439-13

```

```

alignment_scores:
Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

```

alignment_block:

US-09-823-101-1 x US-08-454-439-13 ..

Align seg 1/1 to: US-08-454-439-13 from: 1 to: 576

```

42 ATGCTAACAGACAGACCACTCTTAAGCTTAAGTGAAGGGAAGCGCTGAA 91
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
6  ILeValAlaasnThrValLeuLeuLysAlaArgGluGlyGlyGlyAsp 22
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
92 CAAATAGAGAAAAAGAAAAAGAGAGCTATTAGTGGGGAAGGTAGG 141
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
22 nArgLysGly.LysSerLysLysTrpArgInMetLeu..... 34
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
142 CCAGAGAGCAGCAGAGACCATGACCTTAATATGGAAC...AGAAC 188
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
35 .....GlnPheProHisIleSerGlnCysGluGluLeuArgLe 47
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
189 ATTCTGTAGAGTCACATGCTGCTGAGCCACTGAAATCCAGCTCT 238
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
47 uSerLeuGluArgAspTrpHisSerLeuCysGlu..ArgHisArgIleG 63
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
239 CCAAGTTTCTATATGCAGAG..... 258
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
63 LysArgLeuLeuPheArgGluPheCysAlaThrArgProGluLeuSerArg 79
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
259 .....AAAAAGA 266
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
80 CysValAlaPheLeuAspGlyValAlaGluTrpGluValThrProAspAs 96
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
267 CAGAGAGGAATATGC.....CAAAAT.....AGTCACA 295
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
96 pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
296 GT..... 297
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
113 hArgLysProAspLeuIleProGluValProArgGlnLeuValThrAsnCys 129
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
297 ..... 297
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
130 ThrGlnArgLeuGluGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
298 .....GATTGTGCTGAGGCGGAATTAAGTGGTGTCTTGG 334
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
146 rArgLeuThrHisGluTrpLeuSerValAlaProPhe..AlaAspTrpLeu 162

```

```

335 TTCTTTTATATTCACAAATTTTATAGACACAGAAATACCTTTTAGA 384
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
163 AspSerIleTrpPheAsnArgPheLeuGln...TrpLysTrpLeuGluArg 178
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
385 ATTCTGACTACCAAAATGTGTT..... 409
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
178 gGlnProValThrLysAsnThrPheArgGlnTrpArgValLeuGlyLysG 195
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
410 .....TTTAATTCCTGTGGCAAGTTCAGAAATATGCAATAAAGCAAA 454
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
195 LysGlyPheGlyGluValCysAlaCysGln..... 204
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
455 GATTACTAAAGAAAGAAACTACCTCAAGGTTA.....AAAAACC 495
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
205 .....ValArgAlaThrGlyLysMetCysTrpAlaCysLysLysLe 217
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
496 AAAAAAAAAAAAAAAAAAAGGCGGCGAA 529
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
217 uGluLysLysArgIleLysLysArgLysGlyGlu 228

```

seq_name: /cgn2_6/plodata/2/1aa/PCTUS_COMB.pep:PCT-US94-10487-13

```

seq_documentation_block:
; Sequence 13, Application PC/TUS9410487
; GENERAL INFORMATION:
; APPLICANT: ICOS Corporation
; TITLE OF INVENTION: A Novel G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10487
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,817
; FILING DATE: 31 MAR 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-10487-13

```

alignment_scores:

Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

alignment_block:
US-09-823-101-1 x PCT-US94-10487-13 ..

Align seg 1/1 to: PCT-US94-10487-13 from: 1 to: 576

```

42 ATGCAACAGACGACGACCTCTAGGTTAAAGTGAAGGAGGCTAGAA 91
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
6  lleValAlaAsnThrValLeuLeuValaargGluGlyGlyGlyAs 22
92 CAAATAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
   ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 nArgLysGly.LysSerLysLysTrpArgGlnMetLeu..... 34
142 CCAGAGACGAGGACAGACACATGACCTTAATATGAAAC..AGAAC 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 .....GlnPheProHisIleSerGlnCysGlnGlnLeuArgle 47
169 ATTCTGTGTAGTCCACTCACTGCTGACGTCGACGACGACGACGCT 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
47 userLeuGluArgAspTrpHisSerLeuGln.....ArgHisArgIleG 63
239 CCAGTTTCTATATGACAGAG..... 258
63 LysArgLeuLeuPheArgGlnPheCysAlaThrArgProGlnLeuSerArg 79
259 ..... 266
80 CysValAlaPheLeuAspGlyValAlaGluTrpGluValThrProAspAs 96
267 CAGGAGAGAAATATGC.....CAAAT.....AGTCACA 295
   ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296 GT..... 297
113 hGlnLysProAspLeuLeuProGlnValProArgGlnLeuValThrAsnG 129
297 ..... 297
130 ThrGlnArgLeuGlnGlnGlyProCysLysAspLeuPheGlnLeuTh 146
298 .....GATTTGTCTGAGTGGTGAATATGCTGATCTTGTG 334
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
146 rArgLeuThrHisGluTrpLeuSerValAlaProPhe..AlaAspTrpLeu 162
335 TTCTTTTATATTTCCAAATTTTATAAGACACAGAAATACTTTAAGA 384
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
163 AspSerIleTrpPheAsnArgPheLeuGln..TrpLysTrpLeuGlnArg 178
385 ATTCTGACTACAAATAATGCTT..... 409
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
178 gGlnProValThrLysAsnThrPheArgGlnTrpArgValLeuGlyLysG 195
410 ....TTTAATTCCTGTGCCAAGTTCAGAAATATGCCATATGGAACAA 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 LysLysPheGlyGlyValAlaCysLysGln..... 204
455 GATACTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 495
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
205 .....ValArgAlaThrGlnLysMetTrpAlaCysLysLysLe 217
496 AAAAAAAAAAAAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 529
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
217 uGlnLysLysArgIleLysLysArgLysGlyGlu 228

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-942-686-2

seq_documentation_block:
; Sequence 2, Application US/08942686

```

; Patent No. 6183988
; GENERAL INFORMATION:
; APPLICANT: Bloch, Donald B.
; APPLICANT: Bloch, Kenneth D.
; TITLE OF INVENTION: LEUCOCYTE-SPECIFIC PROTEIN AND GENE, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,686
; FILING DATE: 02-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/027,347
; FILING DATE: 02-OCT-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.4330001/JAG/BJD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-686-2

```

alignment_scores:
Quality: 69.50 Length: 91
Ratio: 1.311 Gaps: 4
Percent Similarity: 58.242 Percent Identity: 26.374

alignment_block:
US-09-823-101-1 x US-08-942-686-2 ..

Align seg 1/1 to: US-08-942-686-2 from: 1 to: 753

```

14 ACCAGAGATCTGAGTGAAGACTCACATTATGGTAAACAGACGAGCTCC 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ThrValAlaIleAlaAsnAsnSerThrLeuGlyLys.....Pr 381
64 TAGAGTTAAGTCAGCGGAAGGCTAGAACAAATAGAAAAAGAGAAAA 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 Lys.....ArgLysArgArgLysL 388
114 AGAGAGGTCATTAGGGGAAAGTAGCCAGAGAGAGAGAGAGAGAGAG 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 LysArgGlyHisGlyTrpSerArgMetArgMetArgGlnLysAsnSer 404
164 ATGACCTTAATATGGAAGAGAGACATCTTGTGTAGGTCACATCATCT 213
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
405 Gln.....GlnAsnAspAsnSerLysAlaAspGlyGlnValVa 417
214 CAGTGCACACTGAAGAAATCCAGCTCCAGTTTCTATATGAGAGAAAA 263
   ::::::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
417 LysSerGlnLysLysAlaAsnValAlaAsnLeuLysAspLeuSerLysIleA 434

```



```

63  LyaArgLeuPheArgGluPheCysAlaThrArgProGluLeuSerArg 79
259  .... AAAAGGA 266
80  CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAsp 96
267  CAGGAGAGAAATATGC.....CAAAAT.....AGTCACA 295
96  pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296  GT..... 297
113  hGlyProAspLeuLeuProGluValProArgGlnLeuValThrAsnCys 129
297  ..... 297
130  ThrGlnArgLeuGlnGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
298  .....GATTTGCTGAGTGTGATTAATGCGTATCTTTG 334
146  rArgLeuThrHisGluTyrLeuSerValAlaProPheAlaAspTyrLeu 162
335  TTCTTTTATATTCCAATTTTATAGACACAGAAATCTTTTAAGA 384
163  AspSerIleTyrPheAsnArgPheLeuGln...TrrLysTrpLeuGlnAr 178
385  ATTTCGACTACAAAATTTGTTT..... 409
178  gGlnProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
410  ....TTTAATTCCTGTGCCAAGTTCAGAAATATGCCATAATGGAACAA 454
195  LylGlyPheGlyGluValLysAlaCysGln..... 204
455  GATTACTAAAGAGAAACTACCTCAAGTTA.....AAAAACC 495
205  ....ValArgAlaThrGlyLysMetTyrAlaCysLysLysLeu 217
496  AAAAAAAAAAAAAAAAAAAGGGGGGGA 529
217  uGluLysLysArgLleLysLysArgLysGlyGlu 228

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-477-451-3

seq documentation block:
: Sequence 3, Application US/08477451
: Patent No. 592865
: GENERAL INFORMATION:
: APPLICANT: Covaccl, Antonello
: TITLE OF INVENTION: Helicobacter Pylori CagI Region
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,451
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0335, 002
: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 510-601-2708
: TELEFAX: 510-655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3174 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-477-451-3

alignment_scores:
  Quality: 68.50      Length: 120
  Ratio: 1.557        Gaps: 4
  Percent Similarity: 36.667      Percent Identity: 24.167

alignment_block:
US-09-823-101-1/rev x US-08-477-451-3 ..
Align seg 1/1 to: US-08-477-451-3 from: 1 to: 3174

537 CCCCTTGTTCCCCCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 488
1100 ProIleGlySerValPhePhePheLeuCysPhePheGlyPhePh 1116
487 TAACCTTGA..... 479
1116 eValIleIleCysProIleArgThrGlnAsnAsnGlnThrGlnAsnPhG 1133
478 .....GGTAGTTTCTTCTTTTAACTTATC..... 455
1133 InLysLysAlaSerPheLeuIleAlaIleIleValAlaSpGlnLeu 1149
454 ....TTGTTTCAT..... 446
1150 SerLeuLeuHisProArgGlnTyrSerGlnThrCysSerAsnArgTyrAr 1166
445 .....TATGCGATATTCTGTAACCTTGCGACGAGAAATTAAGAAC 406
1166 gLeuProProLeuThrLeuThrPValAlaArgLysGlnLysGlnSerPro 1183
405 CAATTTTGTAGTCAGAAATTTCTTAAAGTATTTCTGTGCTTATANA 356
1183 sPhePheCysHisGlnArgPheSerLysTyrProLeuCys..... 1196
355 AATTGGAAATATTAAGAACAAAGAAATCAGCCATTAATTCACACTCAG 306
1197 .....LeuProLeuLys 1200
305 ACAAAATCAC 296
1200 sArgAsnHis 1203

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-748-947A-2

seq documentation block:
: Sequence 2, Application US/08748947A
: Patent No. 5854031
: GENERAL INFORMATION:
: APPLICANT: Jigami et al., Yoshifumi
: TITLE OF INVENTION: MANNOSYL-1-PHOSPHATE TRANSFERASE GENE
: TITLE OF INVENTION: FROM YEAST, AND A PROCESS FOR PRODUCING
: TITLE OF INVENTION: MANNOSYL-1-PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS AND
: TITLE OF INVENTION: PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS BY USE OF THE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 601 Thirteenth Street, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA

```

```

; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,947A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 299509/1995
; FILING DATE: 17-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; REFERENCE/DOCKET NUMBER: 08206/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/783-5070
; TELEFAX: 202/783-2331
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-748-947A-2

```

```

alignment_scores:
  Quality: 67.00      Length: 123
  Ratio: 1.218        Gaps: 6
Percent Similarity: 44.715  Percent Identity: 24.390

```

```

alignment_block:
US-09-823-101-1/rev x US-08-748-947A-2 ..

```

```

Align seq 1/1 to: US-08-748-947A-2 from: 1 to: 446

```

```

353 TTTCGAATATTAAGAACAAAGATCAGCATTAATTCACCACTCAGAC 304
||||:||||| :|||:||||
324 PheSnaSncysGluPheThSerAsn..... 332
303 AAATCAGCTGAGCTATTGTCATATTTCCTGCTCTTTTC... 258
||||:||||| :|||:|||||
333 .....PheGluIleGlyAsnLeuAsnPhenylalrags 343
257 ..TCTGATATAGAAACTG.....GAGAGCTGGAATT 228
||||:||||| :|||:|||||
343 ePrOlaIaTyArgLysPheAsnTyrlleAspGluGlyIle 359
227 TTTCAGTGGCAGCTGAGCATGAGTGACCTACACAGAATGTCGTTC 178
|||:||||| :|||:|||||
360 PheTyItyrPlyStrPse.....AspSe 367
177 CATATAGAGTCATGTGCTCTGCTCTGCTGACCTACCTCCCA 128
||||:||||| :|||:|||||
367 rIleIleIleThrlleGlyLeuSerMetleu.....LeuPrO 380
127 CTAATGACCTCTCTTTTCTCTTTTCTATTTTGTTCAGCCCTTCC 78
|||:||||| :|||:|||||
380 yAsPrLySrlleIleSrlPheGluSnlleGlyPhe..... 391
77 CTCACCTTAACCTTAGAGTCTGTGCTGTGTACATATAGAGTCTCA 28
||||:||||| :|||:|||||
392 .....HisTyAspLysTyAsnAsnSncysProLeuAs 402
27 CTCAGATCTCTGTTCCA 9
||||:||||| :|||:|||||
402 nAsPAsPrlleTyPAsnGln 408

```

```

seq_name: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-09-068-650-5

```

```

seq_documentation_block:
; Sequence 5, Application US/09068650
; Patent No. 6083720
; GENERAL INFORMATION:
; APPLICANT: Chirodoczek Jadwiga
; APPLICANT: Fender Pascal
; TITLE OF INVENTION: Dodecahedral Adenoviral Protein Complex,
; FILE REFERENCE: 3339-368
; CURRENT APPLICATION NUMBER: US/09/068,650
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid peptide
US-09-068-650-5

```

```

alignment_scores:
  Quality: 66.00      Length: 23
  Ratio: 3.667        Gaps: 0
Percent Similarity: 78.261  Percent Identity: 60.870

```

```

alignment_block:
US-09-823-101-1 x US-09-068-650-5 ..

```

```

Align seq 1/1 to: US-09-068-650-5 from: 1 to: 40

```

```

452 CAAGATACATAAAGAGAAAGAACTCAAGCTTAATAAACCAGAAAA 501
:::|||||: :|||:|||||
17 GlusPgluSerLysLysLysLysLysLysLysLysLysLysLys 33
502 AAAAAAAAAAAAAAAAAAG 520
|||||:|||||:|||||
33 slyLysLysLysLysLysLys 39

```

```

seq_name: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-612-126-11

```

```

seq_documentation_block:
; Sequence 11, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; CURRENT FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 11
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human high
; OTHER INFORMATION: molecular weight kininogen light chain amino acids
; OTHER INFORMATION: Glu(448) through Ser(626)
US-09-612-126-11

```

```

alignment_scores:
  Quality: 66.00      Length: 68
  Ratio: 1.737        Gaps: 4
Percent Similarity: 55.882  Percent Identity: 32.353

```


alignment_block:

US-09-823-101-1 x US-09-612-126-11 ..

Align seg 1/1 to: US-09-612-126-11 from: 1 to: 179

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
||||| |||:|||||
23 GLYHISvalleuaspHISGLYHISLys.....HISLysHISGLYHISGL 37
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACATGCA 167
||||| |||:|||||
37 YHISGLYSLysASnLysGLYSLysASnGLYSLysASnGLY. 53
168 CCTTAATATGAAACAGAACATTTCTGTAGGTCACACTCATCTCAGT 217
||||| |||:|||||
54 .....TrrplystHrGLuHIS.....LeuAlaSer 61
218 GCCACTGAAAAATCCAGCTCTCCAGTTTCTATATGCAGAGAAAAAGAGC 267
||||| |||:|||||
62 SerSerGLuAspSerThrThrProSerAlaGLnThrGLnGLYSLYThrGL 78
268 AGGA 271
|||
78 uGly 79
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-612-126-8

seq_documentation_block:

```
; Sequence 8, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT KININOGEN DOMAIN 5
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 8
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human high
; OTHER INFORMATION: molecular weight kininogen light chain amino acids
; OTHER INFORMATION: HIS(441) through Ser(626)
US-09-612-126-8
```

alignment_scores:

Quality:	66.00	Length:	68
Ratio:	1.737	Gaps:	4
Percent Similarity:	55.882	Percent Identity:	32.353

alignment_block:

US-09-823-101-1 x US-09-612-126-8 ..

Align seg 1/1 to: US-09-612-126-8 from: 1 to: 186

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
||||| |||:|||||
30 GLYHISvalleuaspHISGLYHISLys.....HISLysHISGLYHISGL 44
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACATGCA 167
||||| |||:|||||
44 YHISGLYSLysASnLysGLYSLysASnGLYSLysASnGLY. 60
```

168 CCTTAATATGAAACAGAACATTTCTGTAGGTCACACTCATCTCAGT 217

||||| |||:|||||

61TrrplystHrGLuHIS.....LeuAlaSer 68

218 GCCACTGAAAAATCCAGCTCTCCAGTTTCTATATGCAGAGAAAAAGAGC 267

||||| |||:|||||

69 SerSerGLuAspSerThrThrProSerAlaGLnThrGLnGLYSLYThrGL 85

268 AGGA 271

|||

85 uGly 86

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-612-126-1

seq_documentation_block:

```
; Sequence 1, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT KININOGEN DOMAIN 5
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Human
US-09-612-126-1
```

alignment_scores:

Quality:	66.00	Length:	68
Ratio:	1.737	Gaps:	4
Percent Similarity:	55.882	Percent Identity:	32.353

alignment_block:

US-09-823-101-1 x US-09-612-126-1 ..

Align seg 1/1 to: US-09-612-126-1 from: 1 to: 255

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
||||| |||:|||||
99 GLYHISvalleuaspHISGLYHISLys.....HISLysHISGLYHISGL 113
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACATGCA 167
||||| |||:|||||
113 YHISGLYSLysASnLysGLYSLysASnGLYSLysASnGLY. 129
168 CCTTAATATGAAACAGAACATTTCTGTAGGTCACACTCATCTCAGT 217
||||| |||:|||||
130 .....TrrplystHrGLuHIS.....LeuAlaSer 137
218 GCCACTGAAAAATCCAGCTCTCCAGTTTCTATATGCAGAGAAAAAGAGC 267
||||| |||:|||||
138 SerSerGLuAspSerThrThrProSerAlaGLnThrGLnGLYSLYThrGL 154
268 AGGA 271
|||
154 uGly 155
```

